Regularized Synthetic Control Methods:

Advancing Causal Inference in Time Series Econometrics and Observational Studies

Jörg Breitung[a], **Lennart Bolwin**[a], **Justus Töns**[a]

[a]University of Cologne, Chair of Statistics and Econometrics

Abstract

The Synthetic Control (SC) method is a widely used tool for measuring causal treatment effects in observational trials. Typically, the counterfactual of the single treated unit is synthesized using a weighted average of the remaining units in the post-treatment phase. These weights are computed in a data-driven manner and aim to minimize the distance between the treated unit and its counterfactual in the pre-treatment phase. To avoid overfitting the training data and to ensure external validity of the results, the method's developers (Abadie, Diamond, and Hainmueller (ADH)) incorporated the constraint that each weight must be weakly positive, and all weights must up sum to one. Building on the work of [Doudchenko and Imbens, 2016], we propose a generalization that allows for the inclusion of a constant term and negative weights. However, we develop the Regularized Synthetic Controls (REGSC) estimator, an alternative regularization approach that shrinks individual coefficients towards zero and the sum of coefficients towards one. Besides the crucial advantage of a closed form expression, Monte Carlo studies confirm that this regularization method dominates other estimators in data-generating processes with factor structure as was proposed by the inventors of SC. Next, we extend the approach to dynamic contexts and propose a regularized Autoregressive Distributed Lag (ARDL) model for optimal estimation of the counterfactual in time series settings. Again, simulations confirm the new method's potential to enhance the accuracy and robustness of causal effect estimation in time series econometrics and observational studies.

Keywords: *Synthetic Control; Observational Studies; Causal Inference; Regularization, Autoregressive Distributed Lag Models*

List of Acronyms

ADH Abadie, Diamond, and Hainmueller

ARDL Autoregressive Distributed Lag

CV Cross Validation

GDP Gross Domestic Product

DGP Data Generating Process

iid independent and identically distributed

MSE Mean Squared Error

MZ Mincer-Zarnowitz

OLS Ordinary Least Squares

PC Principal Components

REGSC Regularized Synthetic Controls

RMSE Root Mean Squared Error

RSS Residual Sum of Squares

SC Synthetic Control

UK United Kingdom

VAR Vector Autoregression

VARSC Vector Autoregressive Synthetic Control

Contents

List of Figures

1. Introduction

The SC method was developed by Alberto Abadie and colleagues in a series of influential papers ([Abadie and Gardeazabal, 2003], [Abadie et al., 2010], [Abadie et al., 2015]). The method is designed to estimate the causal effect of a treatment in settings with a single treatment unit and a number of potential control units. This is achieved by comparing the treatment unit to a synthesized version of that unit which approximates the counterfactual, i.e. the hypothetical trajectory of the treatment unit in the absence of the treatment. Pre- and post-treatment data are observed for the treatment and control units for the outcome of interest as well as for a set of covariates. Usually, both the number of potential control units J and the amount of pre-treatment periods T_{pre} are of similar sparse magnitude. Consequently, many routinely employed models like Ordinary Least Squares (OLS) are inappropriate as they lack stability and may not even be identified.

This paper is structured as follows: In chapter 2 we present the canonical applications of the SC-method as well as some recent examples and developments. In chapter 3 we examine the trade-off between comprehensibility and statistical optimality in the context of SC. Therefore, we elaborate on the original method as proposed by ADH and propose two alternative estimators suitable to precisely predict the counterfactual. The proposed and the already existing estimators are extensively tested in the subsequent Monte Carlo in which we consider a static factor and a dynamic Vector Autoregression (VAR)-data generating process. Chapter 6 employs our proposed estimates outside the controlled experimental setting and re-estimates the three central SC-applications of ADH. Chapter 7 concludes.

2. Literature Review

Before influential research in the context of SC is presented, we want to introduce a standardized notation. Therefore, without loss of generality, assume that panel data is observed for $j = 0, 1, ..., J$ panel units and that unit $j = 0$ is exposed to treatment at period T_0. The temporal ordering is thus as follows:

$$\underbrace{1, 2, ..., T_0 - 1}_{T_{pre}}, \underbrace{T_0, T_0 + 1, ..., T}_{T_{post}},$$

such that $T_0 - 1$ periods of pre-treatment data (T_{pre}) and $T - (T_0 - 1)$ periods of post-treatment data (T_{post}) is observed.

Canonical Applications

In their canonical 2003 article, [Abadie and Gardeazabal, 2003] evaluate the causal economic effects of conflict using terrorist conflicts in the Basque Country as a comparative case study. Their data consists of $T_{pre} = 15$ $(1955 - 1969)$ periods of pre-treatment and $T_{post} = 28$ $(1970 - 1997)$ periods of post-treatment data for $J = 16$ controls and the single treatment unit indexed by $j = 0$. By constructing a synthesized Basque country that is computed as a weighted average of the remaining regions in Spain that did not experience terrorist conflicts, they invent the SC-method to conduct causal inference in observational settings. The weights are computed such that they optimally match the central variable of interest (Gross Domestic Product (GDP) per capita) as well as a set of time-invariant covariates of the central outcome variable for the treatment unit in the pre-treatment period. Constraining the weights to be weakly positive and to sum up to one provides an easy-to-interpret percentage interpretation and ensures that the synthetic control generalizes well in the post-treatment period. They find that terrorist conflicts caused the per capita GDP of the treatment unit (Basque Country) to decline by about 10% relative to the synthesized control unit.

The estimation of a comprehensive anti-smoking legislation in California in 1988 constitutes another central application of the SC method by [Abadie et al., 2010]. Here, the outcome of interest is per capita smoking in California and 29 U.S. states without such tobacco control programs serve as control units, referred to as Donors. The authors build their estimation on only $T_{pre} = 18$ $(1970-1987)$ pre-treatment years of data and $T_{post} = 13$ $(1988 - 2000)$, indicating the necessity to employ alternatives to the inestimable OLS estimator. [Abadie et al., 2010] reckon that Proposition 99 had a substantial, time-increasing negative effect on per capita cigarette sales by an average of almost 20 packs per person (approximate decline of 25%). Besides presenting the causal treatment effect, the scholars also investigate the statistical significance of their results. By applying a version of the Exact Hypothesis Test proposed by [Fisher, 1935], they find that the probability of experiencing a treatment effect as extreme as observed for California is only 2.6%.

The reunification of East and West Germany correlated strongly with an observable slowdown of GDP per capita growth in West Germany. [Abadie et al., 2015] utilize this natural

experiment as another application of the SC-method. In contrast to other applications of SC, the reunification-dataset is somewhat more wealthy as data is observed for $T_{pre} = 30$ $(1960 - 1989)$, $T_{post} = 14$ $(1990 - 2003)$ years and $J = 16$ donors and West Germany. From 1992 onward, they identify a clear negative average treatment effect of about \$1,600 per capita and year (approximately 8% reduction compared to the 1990 baseline level). In order to examine the trade-off between interpretation and statistical optimality, they sequentially remove donor units from the synthetic control and re-estimate the model. In doing so, they find that the synthetic control heavily relies on one donor (Austria), a peculiarity arising from the specific restrictions (no intercept and percent-like coefficients) of the method.

Further Applications and Developments

[Athey and Imbens, 2016] examine the topic of causal inference in observational studies at a higher level of abstraction and present their often quoted assessment of SC being arguable "the most important innovation in policy innovation in the last 15 years". Besides elaborating on the synthetic control method, their work also includes a careful discussion of related methods for causal inference in observational studies such as regression discontinuity design, differences-in-differences, and particularly in context of high-dimensional settings, machine learning methods.

[Doudchenko and Imbens, 2016] specifically focus on the identifying assumptions of SC. Besides the thorough treatment of the relationship between the amount of explanatory variables J and pre-treatment observations T_{pre}, they focus in great detail on the four prevailing restrictions on the intercept and the weights, namely the no-intercept assumption, the adding-up and non-negativity assumption as well as the potential restriction of constant weights. Their main recommendation is to leave the restrictions aside and to opt for an elastic-net regression that ensures external validity by means of regularization. They apply the elastic together with three other proposal models to three core applications of SC and obtain comparable results for the different models.

The SC-method is also widely used in contemporary research: For example, [Born et al., 2019] apply the method to quantify the economic cost of nationalism in context of the Brexit referendum vote and find that the so-called "doppelganger gap", i.e. the difference between actual and synthesized GDP-trajectory of the United Kingdom (UK) ranged between 1.7 and 2.5% after UK's citizens voted against remaining in the EU. To disentangle their estimated treatment effect, the scholars proceed in two steps: First, by disassembling GDP into its components, they find that consumption and investment are the main drivers of the decline. Second, they estimate an expectation-adjusted VAR-model to explicitly account for anticipation and uncertainty.

Another recent synthetic control article was written by [Muhlbach and Nielsen, 2019]. The authors study the causal effect of the relocation of the US embassy from Tel Aviv to Jerusalem on the weekly number of conflicts in Israel and Palestine. However, instead of relying on the conventional SC method or related linear models, they opt for a random

forest regression. They prove the asymptotic unbiasedness of the estimate as well as its consistency and conclude that their approach is not dominated by the remaining comparison models like the SC. As expected, the relocation of the embassy caused an increase of conflicts. By conducting various inferences tests, the scholars find statistical significance at the 1% level.

[Harvey and Thiele, 2020] pursue a multivariate structural time series approach to estimate the counterfactual for the treatment unit. Importantly, they argue that the decision for or against a potential donor should be made on the basis of stationarity tests and show such tests are applicable in small samples. They re-estimate two of the core SC applications (tobacco control and German reunification) and conduct the KPSS-test for the difference between the treatment series and each potential donor to analyse stationary patterns. If the test rejects the null of stationarity around a deterministic trends, this serves as evidence against a balanced growth model and the potential donor should not be included in the counterfactual.

An incomplete list of recent SC application also includes [Cho, 2020] and [Cunningham, 2021]. Cho quantifies the impact of non-pharmaceutical interventions during the COVID-19 outbreak in Sweden and obtains robust indications for the adverse public health effects of tentative intervention policies during the COVID-19 pandemic. Cunningham studies the effect of incarceration in Texas on drug markets. Besides identifying only moderate effects of Texas doubling the state's prison capacities on the drug market, Cunningham has a salient point on the practical use of SC: "Authors using synthetic control must do more than merely run the synth command when doing comparative case studies. They must find the exact p-values through placebo-based inference, check for the quality of the pre-treatment fit, investigate the balance of the covariates used for matching, and check for the validity of the model through placebo estimation [...]."

[Ben-Michael et al., 2021] connect to the commonly violated assumption of a perfect pre-treatment fit of the original SC method: they introduce the augmented synthetic control method that accounts for a potential bias of the SC-estimator due to imperfect pre-treatment fit. Simplified, their model uses a ridge penalty to improve the pre-treatment fit and penalizes extrapolation from the convex hull of the donors. In some applications, interventions are subject to staggered adoption, i.e. multiple panel units adopting a given given treatment at different times. [Ben-Michael et al., 2021] generalize the original SC-method to such scenarios by proposing a "partially pooled" SC that allows for unit-level intercepts and covariates.

Last but not least, [Abadie and L'Hour, 2021] propose the so-called penalized synthetic control estimator that is closely related to staggered adoption. The development is motivated by the observation that the original SC method is increasingly applied to settings of multiple treated units. In such contexts, the overall treatment effect can be approached through two distinct methodologies. The first entails aggregating all treatment units and subsequently calculating the synthetic control, whereas the second involves the computation of synthetic controls for each unit, followed by the aggregation of these individual

synthetic controls. Specifically when the matching variables for the treatment units fall inside the convex hull of the donor pool, nonuniqueness of the weights is likely to be a concern. To solve this issue, the authors provide a regularization that trades off the competing objectives of minimizing covariate discrepancy between each donor and its specific synthetic control and the objective of minimizing covariate discrepancy between each donor and the aggregated synthetic control.

Hypothesis Tesing

The question of treatment effect significance arises naturally subsequent to the construction of the synthetic control. ADH propose a model-invariant non-parametric inference procedure that is based on the Exact Hypothesis Test proposed by [Fisher, 1935]. ADH consider permutations in region (i.e. panel unit) and time. Region permutations estimate the treatment effect for each panel unit $j \in \{0, ..., J\}$.[1] This procedure provides the researcher with the empirical $(J+1)$-observational distribution of the treatment. Next, the estimated treatment effect of the truly treated unit can be compared to the J placebo-treatment effects of the donors. Given the estimated treatment effect for the truly treated unit is large, the null hypothesis of no treatment effect can be rejected at the significance level of one minus the percentile of treatment effect in the empirical distribution.[2] Time permutations consider only the truly treated unit and permute the treatment to time periods prior to the true treatment date T_0. Given that $T_0 > J$, this approach can increase the sensitivity of the test, since the theoretically feasible significance threshold of region permutation tests is determined by $\frac{1}{J}$. For both, region and time permutations, ADH condense the estimated treatment effects into a precision metric like the post-treatment Mean Squared Error (MSE). Outlier in the donor pool can cause problems: Suppose for instance, that one donor region is very different from the rest such that it falls outside the convex hull of the remaining donors.[3] Since the outlier itself cannot be synthesized precisely by the donor pool, pre- and post-treatment MSE are expected to be large. Consequently, the permutation test will be unreasonably conservative and ADH propose to exclude regions that are hard to predict, i.e. who have a pre-treatment MSE that exceeds the MSE of the truly treated unit to a great extent. Figure 1 visualizes the exclusion procedure in the tobacco control application of ADH.

[1] Note that it is necessary to exclude the truly treated unit from donor pool as otherwise, it will contaminate the synthetic control

[2] For instance, let $J = 99$ such that treatment effects for 100 panel units can be computed. As long as the estimated treatment effect of the truly treated units belongs to the 95 most extreme effects (95th percentile or higher), the permutation test rejects the null hypothesis of no treatment effect at least at 5 percent.

[3] Note, that this circumstance does not cause problems for the truly treated region and its synthesized counterfactual as we expect the SC-algorithm to assign a near zero weight to such an outlier.

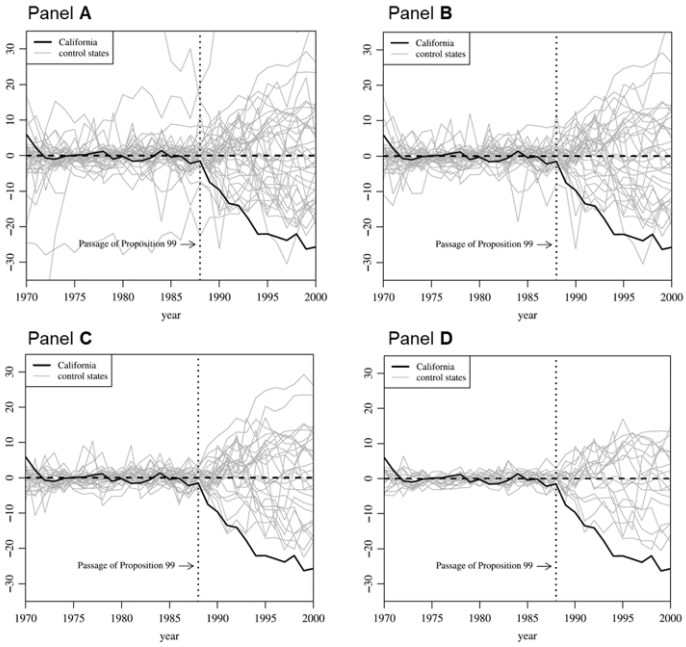

Figure 1. Region Exclusion Procedure of ADH

The vertical axis indicates the gap between observed and estimated per capita cigarette sales, the bold line represents the truly treated region (California). Some regions have both a poor pre- and post-treatment fit. Since the treatment significance should not be artificially driven by regions with poor fit, ADH successively remove regions with a large pre-treatment MSE relative to California. Panel B excludes regions with a MSE that is more than 20 times as large the MSE of California, Panel C lowers the cutoff to five times California's MSE and Panel D to two times the MSE. In the last scenario, only 19 regions are left and California is the one with the most extreme treatment effect. The authors therefore conclude that the treatment is statistically significant with a (permutation) p-value of 5,3% $\left(\frac{1}{19}\right)$. One way to bypass the inefficient sample reduction procedure is to look at the distribution of the ratios of pre- and post-treatment MSE. By scaling the post-treatment fit by the pre-treatment fit, regions with a poor fit are implicitly controlled for. In the tobacco control application, California is the region with the highest pre- to post-treatment MSE ratio among all 39 regions, translating into a p-value of 2.6% $\left(\frac{1}{39}\right)$. The end-of-sample test by [Andrews, 2003] for structural breaks is another frequently applied econometric test to assess treatment effect significance. In comparison to the region permutation test by ADH, it does not rely on the assumption of random treatment. This distinction is crucial as 1) the random treatment assumption is unrealistic in many SC applications and 2) it does not require estimating the model for all units. Though the specific test statistic of the test is case specific, a commonly employed variant is based on the weighted sum of squared residuals. Under the null hypothesis of no structural break, this test statistic typically follows a known distribution like the standard normal and deviations indicate treatment effect significance. The test can be considered as a

generalization of the F test by [Chow, 1960], allowing for more general error distributions than the one of the linear regression model.

[Chernozhukov et al., 2021] take a similar perspective as [Andrews, 2003] and interpret the causal inference problem as a counterfactual prediction task. Their test builds on the stochastic process of the time-specific difference between observed and synthesized outcome of the treatment unit. Under the null of no treatment effect, the pre- and post-treatment distribution of this processes should be indistinguishable. [Chernozhukov et al., 2021] admit their tests relation to [Andrews, 2003] end of sample test but highly important differences: First, they prove the high small sample performance of their test while Andrew's test relies on asymptotic analyses. Second, Andrews test assumes correct model specification, the test by Chernozhukov and colleagues is also valid under misspecification. Lastly, while Andrew's test is designed for low-dimensional models, their test generalizes well to high-dimensional models like tree-based estimators or neural nets.

Even in case the aforementioned tests hint toward general treatment effect significance, beyond some maximum forecast horizon, forecasts may become uninformative and volatile. [Breitung and Knüppel, 2021] develop tests for the null hypothesis that forecasts from a survey of forecasters or from an estimated (parametric) model provide no more information than the unconditional mean of the series, i.e. that they become uninformative. In context of SC, due to the fact that contemporaneous donor values are available over the entire post-treatment period, it is expected, that the null of uninformative forecasts is rejected at a significantly larger margin. Independent of the specific test that is employed, prediction intervals should be computed to account for general forecasts uncertainty. For example, [Born et al., 2019] define a prediction interval as plus minus one standard deviation of the pre-treatment prediction. If a normal approximation of the residuals seems reasonable, the 95%-prediction interval would be comprised by plus minus 1.96 standard deviations of the pre-treatment prediction. Commonly, prediction intervals increase with the forecasts horizon to account for the fact that more uncertainty is associated with the forecast, the further ahead the forecast. Therefore, the naive standard deviation of a h-step forecast is usually given by $\hat{\sigma}\sqrt{h}$. Lastly, by means of bootstrapping of rejection sampling, the prediction interval could also be simulated. To bootstrap the prediction interval, one could draw e.g. $1,000$ observations with replacement from the data at hand, compute $1,000$ pre-treatment prediction standard deviations and define the 95%-bootstrap prediction interval boundaries by the 25th and 975th largest standard deviation. To apply a rejection sampling approach, similar to Bayesian posterior sampling, one would generate a sample of, e.g. $1,000$ standard deviations obtained by the accept-and-reject cycle of the rejection sampling approach and define the prediction interval analogously to the bootstrap interval (see e.g. [Diebold, 2017], [Kirchgässner et al., 2012] and [Koop, 2003]).

3. Theory

In this chapter, we propose alternative SC-estimators to assess the magnitude of treatment effects in observational settings. We build on the already introduced notation, such that $J+1$ panel units indexed by $j = 0, 1, ..., J$ are observed over a time horizon of T periods. Unit $j = 0$ is exposed to the treatment at period $t = T_0$ with $1 < T_0 < T$. Further, it is assumed that no treatment anticipation and contamination (i.e., no spillovers in time and space) are present. The former would be the case if the treatment affects unit $j = 0$ before T_0, the latter describes the case where some of the supposedly untreated units $j = 1, ..., J$ are contaminated as they are affected by the treatment. To contextualize these assumptions, [Abadie et al., 2010] argue that in the presence of anticipation effects, T_0 could be shifted back in time until the no-anticipation assumption seem plausible. If panel units in the donor pool are affected by the treatment (contamination) as it is likely in the Brexit-application, those units could be removed from the sample prior to the estimation. Our goal is to evaluate the causal effect of the treatment, the specific functional form of it remains unspecified though. This is possible because the main goal of the SC-estimation lies in the precise estimation of the counterfactual. As the treatment scenario is empirically observable, it is not necessary to specify the specific functional form of the it.

3.1. ADH Case

The estimation task can be constituted by the potential outcome framework as introduced by [Neyman, 1923] and elaborated by [Rubin, 1974]. Let $y^I_{j,t}$ be the (potential) outcome for unit j at point t in the presence of the intervention. Likewise, let $y^N_{j,t}$ be the (potential) outcome for j at point t in the absence of the intervention. ADH define the treatment effect of the intervention as

$$\delta_{j,t} = y^I_{j,t} - y^N_{j,t}$$

and introduce the indicator variable $D_{j,t}$ that takes on the value 1 if unit j is treated at period t and the value 0 otherwise. Given the assumed absence of anticipation and contamination, the following outcome is observed

$$y_{j,t} + D_{j,t}\delta_{j,t} = \begin{cases} y^N_{j,t} & (\text{if } j = 0 \text{ and } t < T_0) \text{ or } j \geq 1, \\ y^N_{j,t} + \delta_{j,t} & \text{if } j = 0 \text{ and } t \geq T_0. \end{cases}$$

The goal to estimate the causal treatment effect $(\delta_{0,T_0}, ..., \delta_{0,T})$ therefore boils down to the estimation of the counterfactuals of unit $j = 0$ in the post-treatment phase $(y_{0,T_0}, ..., y_{0,T})$, i.e. on what trajectory would unit $j = 0$ have been, was there no intervention. The basic idea of ADH is to estimate these counterfactuals as a weighted average of the donor outcomes using a data-driven approach to compute the weights. Intuitively, the weights are computed such that they optimally predict the outcomes and a set of time-invariant

explanatory variables for the treatment unit in the pre-intervention phase, conditional on having a percentage interpretation. Thus, for the computation of the weights, we focus exclusively on the pre-intervention time periods $t \in \{1, 2, ..., T_0 - 1\}$. Subsequently, the counterfactuals are extrapolated by applying the calculated weights to the post-intervention time periods $t \in \{T_0, T_0 + 1, ..., T\}$.

Let $Y_j = (y_{j,1}, ..., y_{j,T_0-1})'$ be the vector of observed pre-intervention outcomes for unit j.[4] To distinguish treatment unit and donors, ADH collect the treatment unit in the $((T_0 - 1) \times 1)$-vector Y_0 and row-bind all donor unit vectors into the $((T_0 - 1) \times J)$-matrix Y_1. Moreover, a set of K time-invariant covariates of Y_j is observed for all panel units.[5] Therefore, let X_0 denote the $(K \times 1)$-vector of covariates for Y_0 and let X_1 denote the $(K \times J)$-matrix of explanatory variables for Y_1. To estimate the causal effect of the treatment, the SC-estimator estimates the counterfactuals $(\widehat{y}_{0,1}, ..., \widehat{y}_{0,T_0}, ..., \widehat{y}_{0,T})$ of the single treated unit for the pre- and post-intervation phase as

$$\widehat{y}_{0,t} = \sum_{j=1}^{J} \widehat{w}_j y_{j,t}^N \; \forall \; t \in \{1, ..., T\}$$

The weights $(\widehat{w}_1, ..., \widehat{w}_J)$ are constraint such that $\widehat{w_j} \geq 0 \; \forall \; j$ and $\sum_{j=1}^{J} \widehat{w_j} = 1$. It is worth noting that this constraint requires the counterfactuals to belong to the convex hull of the donors as otherwise, $\widehat{Y_0}$ will never match its true counterpart. [Abadie et al., 2010] argue that "the magnitude of discrepancy" should be calculated in advance of each SC-application. If the researcher finds that the pre-intervention values of Y_0 fall outside the convex hull of the donors, the usage of SC is not recommended. Formally, $(\widehat{w}_1, ..., \widehat{w}_J)$ is the solution of the following nested optimization problem:

$$\widehat{w}(v) = \arg\min_{w} \sum_{k=1}^{K} v_k \left(x_{0,k} - \sum_{j=1}^{J} w_j x_{j,k} \right)^2$$

with v being an arbitrary positive definite vector of dimension $(K \times 1)$ which solve the second optimization problem:

$$\widehat{v} = \arg\min_{v} \sum_{t=1}^{T_0-1} \left(y_{0,t} - \sum_{j=1}^{J} \widehat{w}_j(v) y_{j,k} \right)^2$$

Afterwards, the causal effect of the intervention $\delta_{j,t}$ can be quantified at each time point after the intervention $t \in \{T_0, T_0 + 1, ..., T_1\}$ as the gap between observed $(y_{0,t}^N + \delta_{j,t})$ and predicted outcome $(\widehat{y}_{0,t}^N)$.

This two-step estimation procedure serves two crucial purposes: $\widehat{v}$ measures the relative

[4] For instance, in the canonical example of [Abadie and Gardeazabal, 2003], Y_0 would be the vector of GDPs for Great Britain until the Brexit referendum.

[5] In the already mentioned Brexit-example, natural predictors of GDP are its components consumption, investment, government spending and net exports.

importance of the K variables in X_1 to explain X_0. In contrast, the weighting vector $\widehat{w}(v)$ quantifies the relative importance of each unit in the donor pool. Summarizing the key concept of ADH, the SC-method ensures that the synthesized treatment unit is as similar as possible to the actual treatment unit with respect to the quantity of interest and a set of potential explanatory variables in the pre-treatment period. Especially in the canonical examples of SC, the quantity of interest (e.g. GDP) and the explanatory variables (e.g. consumption, investment, government spending and net exports) are interconnected by construction. Thus, observing that the SC-estimator was capable of approximating both targets significantly enhanced the methods credibility. If the explanatory variables are omitted, the SC-algorithm reduces to an OLS estimation, constraint to have no constant and weakly positive coefficients that sum up to one.

3.2. Simple Static Extension

To provide an intuitive introduction to our proposed extensions, we first consider the most simple scenario of one treatment unit $j = 0$ and two donor units $j = 1, 2$. We consider a setting where only the outcome series (e.g. GDP) and no further covariates (e.g. consumption, investment etc.) are observed. It is assumed that before $t = T_0$ the units have a joint distribution of the form[6]

$$Y = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma) \text{ for } t < T_0.$$

with $\mu = (\mu_0, \mu_1, \mu_2)'$ and the positive definite covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma_{12}'} \\ \boldsymbol{\sigma_{12}} & \mathbf{\Sigma_2} \end{pmatrix}.$$

σ_0^2 denotes the variance of y_0, $\mathbf{\Sigma_2}$ is a (2×2) covariance matrix of the vector $(y_1, y_2)'$ and $\boldsymbol{\sigma_{12}}$ is a (2×1) vector with elements $cov(y_0, y_1)$ and $cov(y_0, y_2)$.

Disregarding any constraints, we are interested in deriving the best unbiased forecast of y_0 given the controls y_1 and y_2 which is obtained as

$$\begin{aligned} \widehat{y}_0^N &= \mu_0 + w_1^{OLS}(y_1 - \mu_1) + w_2^{OLS}(y_2 - \mu_2) \\ &= \mu^* + w_1^{OLS} y_1 + w_2^{OLS} y_2, \end{aligned}$$

where $\mu^* = \mu_0 - w_1^{OLS}\mu_1 - w_2^{OLS}\mu_2$. This forecast can be directly estimated by an unrestricted OLS regression of y_0 on y_1 and y_1. However, the result implies that there is no inherent reason to impose the restrictions that $w_1^{OLS}, w_2^{OLS} \geq 0$ and $w_1^{OLS} + w_2^{OLS} = 1$. Furthermore, we argue that the construction of SC should include a constant term, as

[6] For the ease of exposition we suppress the time index t as in this section we neglect any dynamic effects which will be considered in the next section.

otherwise the estimated counterfactual may have a mean outside the convex hull of the donor means. See also [Doudchenko and Imbens, 2016] for a careful discussion of these restrictions.

For illustrative reasons, assume that

$$Y \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right).$$

For this example the unrestricted optimal weights for the counterfactual result as $w_1^{OLS} = -0.1333$, $w_2^{OLS} = 0.4667$ and $\mu^* = \mu_0 - w_1^{OLS} \cdot \mu_1 - w_2^{OLS} \cdot \mu_2 = 0.6667$.[7] Note that w_1^{OLS} is negative even though all bivariate correlations between the units are positive. One may argue that this result does not make much sense as the economic interpretation of y_1 entering the counterfactual $\widehat{y}_0^N$ with a negative sign is unclear. This demonstrates the trade-off between optimality in a statistical sense and the economic interpretation of the solution. What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination $\widetilde{y}_0^N = 0.2y_1 + 0.8y_2$. The important difference lies in the variance of these estimates. For our example we obtain

$$var(y_0 - \widehat{y}_0^N) = 0.8267$$
$$var(y_0 - \widetilde{y}_0^N) = 1.1600.$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of y_0. This is possible as $(w_1, w_2) = (0, 0)$ is not included in the restricted parameter space.

In microeconometric settings it is usually assumed that the units in the treatment group and units in the control group are uncorrelated. In such cases the construction of a SC is unpromising as the dependency between treatment unit and donors is the core condition for a plausible estimation of the counterfactual. If no such relationship exists, the optimal estimate boils down to $\widehat{y}_0^N = \mu_0$ and, therefore, it does not make sense to involve a SC. In macroeconomic applications however, the variables in the treatment and control group (e.g. GDP) are typically correlated and it is therefore important to model this relationship. As the simple scenario with only two panel units in the donor pool is unrealistic in practice, we now move to the general static case with $J + 1$ panel units.

3.3. General Static Extension

In empirical macroeconomic practice, the observed time series are typically low- frequency, i. e. the quantities of interest are measured at monthly, quarterly or even annual

[7] The derivation of the employed estimators is postponed to the appendix.

intervals. Thus, the number of pre-intervention time periods (T_0-1) is typically small and may even be smaller than the number of units in the donor pool J. In such scenarios, the unrestricted OLS estimate may face issues of instability or, in the case of $T_0 - 1 < J$, due to singularity, it may not even be identified. So see this, let us now consider the statistical properties of the corresponding least-squares estimator for an arbitrary J:

$$y_{0,t} = \mu^* + w_1\ y_{1,t} + w_2\ y_{2,t} + ... + w_J\ y_{J,t} \text{ for } t = 1, 2, ..., T_0 - 1.$$

From standard results on least-squares regressions it follows that for fixed J and $(T_0-1) \to \infty$ the OLS estimator $\widehat{w} = (\widehat{w}_1, ..., \widehat{w}_J)$ is unbiased and converges in probability to the MSE optimal weights w. In empirical practice, we typically have a large number of donors candidates such that J may be of similar magnitude than $(T_0 - 1)$. In this case, $\frac{J}{T_0-1}$ is substantially larger than zero and, therefore, some regularization is required. Indeed, as shown in the next proposition, the OLS estimator is inconsistent in such cases:

Proposition 1 *Let* $Y_t = (y^N_{0,t}, y_{1,t}, \ldots, y_{J,t})'$, $\widehat{y}^N_{0,t} = \widehat{w}' x_t$, $x_t = (y_{1,t}, \ldots, y_{J,t})'$ *and* $\widehat{w}$ *denote the OLS estimator of* w *from above. If* y_t *are independent draws from* $Y \sim \mathcal{N}(\mu, \Sigma)$ *for* $t = 1, \ldots, T_0, \ldots, T$ *then for* $T_0 - 1 \to \infty$ *and* $\frac{J}{T_0-1} \to c > 0$ *it follows that* $\widehat{Y}^N_{0,t} - Y^N_{0,t}$ *is asymptotically distributed as* $\mathcal{N}(0, c)$ *for* $t > T_0$.

It is important to note that the OLS estimator does not converge if both the number of pre-treatment observations and the number of regressors tend to infinity at the same rate. Similar results were obtained by [Bekker, 1994] who considers the asymptotic distribution of $\widehat{w}$. Our result is simpler as we consider some particular linear combination given by $\widehat{w}' x_t$ where $t > T_0 - 1$. In this case the distribution does not depend on the covariance matrix Ω.

The issues of external validity and overfitting are closely related to the aspect of identification. Especially when employing non-parametric statistical learning methods, it is simple to achieve a high in-sample (pre-treatment) fit. The crucial part when dealing with forecasts is that the observed in-sample patterns generalize well outside the verifiable horizon (post-treatment). ADH solve this issue by restricting the weights to be non-negative and to sum up to one. Besides preventing the model from overfitting, the percent restriction guarantees the existence of unique weights, especially when dealing with a small number of pre-treatment periods. Regularized regressions constitute another model family that is capable of balancing the trade-off between under- and overfitting.

In this context [Doudchenko and Imbens, 2016] suggest employing an elastic net regression to regularize the donor weights. It solves the following objective function:

$$Q(w, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0-1} \underbrace{\left(y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j y_{j,t} \right)^2}_{RSS} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2 \right)}_{Ridge} + \lambda_2 \underbrace{\left(\sum_{j=1}^{J} |w_j| \right)}_{Lasso}$$

The L_2-norm (Ridge-Penalty) is a continuous shrinkage method, that shrinks the coefficients towards zero without performing variable selection in the sense that certain coefficients are set exactly to zero ([Hoerl and Kennard, 1970]). However, it has the appealing feature that its estimation only involves the addition of a diagonal matrix to the Residual Sum of Squares (RSS). Therefore, the objective function keeps an explicit closed form solution which is particularly appealing if the sample is small.

In contrast, the L_1-norm (Lasso-Penalty) as proposed by [Tibshirani, 1996] penalizes the sum of the absolute values of the coefficients. The nature of the penalty term causes this regularization to perform both, continuous shrinkage and automatic variable selection. As a consequence, the argmin vector of the objective function typically contains many entries that are exactly zero which makes the resulting model sparse and easier to interpret. However, since the absolute value function is not continuously differentiable, the Lasso has no closed form solution. Consequently, the minimum of the objective function has to be approximated, which is typically done via numerical optimization techniques like cyclical coordinate descent algorithm (see for example [Friedman et al., 2010]). The shrinkage parameters λ_1 and λ_2 can be selected through k-fold Cross Validation (CV). This involves storing combinations of λ_1 and λ_2 that minimize the objective function across k validation sets. The average value of these hyperparameters is then computed to make the final choice.

We propose a different regularization that we call the regularized synthetic control estimator. This estimator augments the OLS objective function by a Ridge penalty and a simple "inverse"-Ridge that shrinks the coefficient sum towards one. The objective function has the following form:

$$Q(w, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0-1} \underbrace{\left(y_{0,t} - \mu^* - \sum_{j=1}^{J} w_j y_{j,t} \right)^2}_{RSS} + \lambda_1 \underbrace{\left(\sum_{j=1}^{J} w_j^2 \right)}_{Ridge} + \lambda_2 \underbrace{\left(1 - \sum_{j=1}^{J} w_j \right)^2}_{"inverse"\text{-}Ridge}$$

Due to the individual shrinkage to zero (Ridge) and the joint shrinkage to one (inverse Ridge), this regularization is closely related to original SC estimator but in contrast to the elastic net, it is flexible enough to produce non-zero weights that are directly interpretable. Moreover, as it does not involve approximating the gradient of the absolute value function, it has the following closed form solution:

$$\widehat{w}_{\lambda_1,\lambda_2} = (Y_1' Y_1 + \lambda_1 I_J + \lambda_2 \mathbf{1_J 1_J'})^{-1} (Y_1' Y_0 + \lambda_2 \mathbf{1_J}).$$

I_J depicts a J-dimensional vector of ones, $\mathbf{1_J 1_J'}$ an all-ones matrix of dimension J. Consistent with the notation of chapter 3.1, Y_1 is a $(T_0-1)\times J$ matrix that stacks all pre-treatment donor observations for $t = 1, ..., T_0 - 1$ and $j = 1, ..., J$. Analogously, Y_0 is a $(T_0 - 1) \times 1$ vector that stacks the pre-treatement time series observations of the treatment unit. To omit the constant from the penalty, a vector of ones should be joined to Y_1 from the left

and the first element of I_J respectively $\mathbf{1_J 1'_J}$ block be set to zero. Alternatively, Y_0 and Y_1 could be demeaned prior to an intercept-free estimation. In the appendix, we show for $\lambda_1 \to \infty$ and $\lambda_1/\lambda_2 \to c$ the weights converge to $1/(n+c)$, which seems to be a more reasonable target than shrinking towards zero as done by the elastic net. Similar to the case of the elastic net, the shrinkage parameters λ_1 and λ_2 can by chosen by cross validation, where our experience suggests that optimizing subject to the restriction $\lambda_1 \approx 10,000 \cdot \lambda_2$ reduces computation time and already produces reasonable estimates.

The combination of a closed form solution and tuneable hyperparameters make the REGSC-method highly appropriate for the low-frequency macroeconomic context of SC: It is able to produce weights that are interpretable, flexible and efficiently estimated in small samples. These characteristics are empirically verified in the subsequent simulation study. Besides the proposed regularization, we also implemented a numerically solvable combination of the Lasso- and the "inverse"-Ridge-penalty. In large data sets of a least $1,000$ observations, this alternative was competitive to the elastic net and the proposed REGSC-estimator. However, as we are looking for an estimator with robust small sample properties, the Lasso-"inverse"-Ridge estimator is omitted from the further analysis.

3.4. General Dynamic Extension

TBD: When modeling macro time series it is often assumed that the $(J+1) \times 1$ vector of time series $y_t = (Y_{0,t}, ..., Y_{J,t})'$ can be represented by a VAR model of the following form:

4. Simulation

In this chapter, we empirically test the performance of our proposed and the already existing estimators for SC in different DGP. Independent of the specific features of the data in terms of pre- and post-treatment period length and the prevailing time series structures, we proceed as follows: We simulate $T_{pre} = T_0 - 1$ periods of pre-treatment and $T_{post} = T - (T_0 - 1)$ periods of post-treatment data for the single treated unit and the J donor units. Each estimator's main goal is to grasp the consistent patterns before treatment and accurately extend these into the time after treatment. Said differently, the pre-treatment phase depicts the training set of the models and the post-treatment the validation set. To root the simulation framework as close as possible to real-world SC applications, we define T_{pre} and T_{post} such that their range is comparable to low-frequency macroeconomic settings, i.e. $T_{pre} \in \{20, 50, 100\}$ and $T_{post} \in \{10, 20, 30\}$. Furthermore, we consider two types of DGP, a static factor process and a dynamic VAR-process that is inspired by real GDP processes of the G20 countries.

4.1. Static Data Generating Processes

4.1.1. Set up

In their SC application of estimating the causal effect of California's proposition 99, [Abadie et al., 2010] suppose that the (potential) outcome $Y_{i,t}^N$ follows a linear factor model of the form

$$Y_{i,t}^N = \alpha_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}.$$

α_t denotes an unknown panel-invariant intercept, Z_i is a vector of observed panel-specific covariates, θ_t is a vector of unknown parameters, λ_t is a vector of unknown common factors and μ_i are panel-specific unknown factor loadings. The unobserved shocks ϵ_{it} have zero mean at the panel level. For this specific setting, [Abadie et al., 2010] show that "[...] the bias of the SC-estimator can be bounded by a function that goes to zero as the number of pre-treatment periods increases." Further the number of donor units has to be fixed. The fact, that α_t is panel-invariant seems minor at first glance. However, as the SC-estimator does neither contain an intercept nor does it allow for extrapolation outside the convex hull of the donor pool, the unbiasedness of the estimator directly depends on the distribution of the intercepts. In a slightly more realistic data-generating scenario, the intercepts do not follow a degenerate point distribution with $P(X = 0) = 1$ but are drawn from a symmetric distribution centered around the origin like the standard normal.

[Ferman, 2021] considers a de-meaned scenario without additional covariates. Here, we follow the basic set-up of Ferman and generate data according to a similar factor model. However, we consider it more realistic to add a time-invariant and panel-specific intercept to the (potential) outcome instead of analyzing a de-meanded DGP. Our representation

of the counterfactuals therefore boils down to

$$Y_{i,t}^{N} = \alpha_i + \lambda_t \mu_i + \epsilon_{it}.$$

In this simplified setting, the counterfactual is given by the composition of the unknown panel-specific factor loadings μ_i and the F unknown common factors $\lambda_t = (\lambda_{1,t}, ..., \lambda_{F,t})$ plus intercept α_i and idiosyncratic shocks ϵ_{it}. For the sake of simplicity, Ferman considers a scenario with only two common factors, $\lambda_{1,t}$ and $\lambda_{2,t}$. We proceed analogously and generate data such that the (potential) outcome of the treated unit and the first half of the donor pool load exclusively with loading one on the first factor, the remaining donors load exclusively with loading one on the second factor. Therefore μ_i is a (2×1)-vector with the first (second) entry being one and the second (first) entry being zero for the first (second) half of the donor pool. Further, the random variables $\alpha_i, \lambda_{1,t}, \lambda_{2,t}$ and ϵ_{it} are realizations of a standard Gaussian white noise process $(\mathcal{N}(0,1))$. The following figure exemplifies the functionality of the DGP with $T_{pre} = 20$ and $T_{post} = 10$ and a constant treatment effect of $\delta_{0,t} = 10$ for $t \geq T_0$.

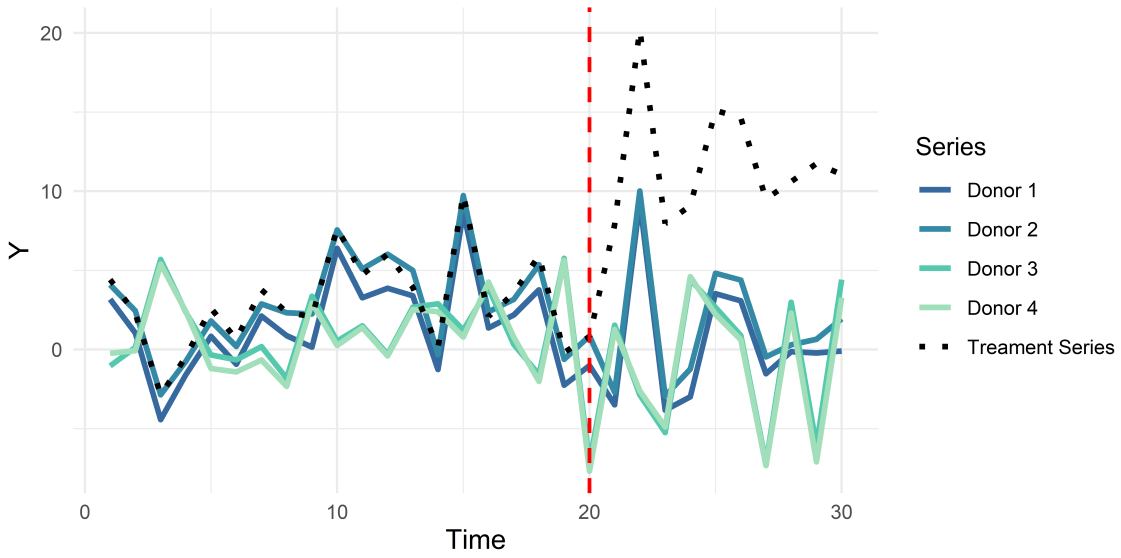

Figure 2. Example Factor-DGP

To make the factor structure more tangible, we scaled the factor variances by 10^1 and the error variance by 10^{-1}. The generated data exhibits a clearly observable factor structure: The treatment unit and the first half of the donors (Donor 1 and 2) as well as the second half of the donors (Donor 3 and 4) share a common factor. Thus, the objective of each employed method is to recover the true factor structure, i. e. irrelevant of the size of donor pool to weight only the first $\frac{J}{2}$ donors positively. Further, we see that each series possesses an own intercept. Yet in this specific example only, the intercept variation is dominated by the factor variation due to the scaling of the variances.

4.1.2. Employed Models

For the static factor DGP, we employ five models:

Synthetic Control (SC)

The first model is the ordinary SC method without additional covariates. Therefore, this method is equivalent to a restricted OLS regression that regresses the treatment series on the donors series given the constraint of no intercept and non-negative coefficients that sum up to one. The belief for this model is that the untuneable restrictions prevent it from overfitting the pre-treatment data but that this inelasticity comes at the costs of a reduced performance in terms of prediction and forecasting.

Ordinary Least Squares (OLS)

The second model is a usual least squares regression that regresses the treatment series on the donor series. The belief for this model is that it starts overfitting the pre-treatment data as J grows large as it exhibits no regularization opportunities. Further for $J > T_0 - 1$, it is unable to provide any prediction or forecast.

Elastic Net Regression (NET)

The third model we consider is the aforementioned elastic net as proposed by [Doudchenko and Imbens, 2016] in the context of SC. Due to its flexible hyperparameters tuned via pre-treatment CV, we expect this model to perform reasonable well over the entire observation window (pre- and post-treatment). For the sake of simplicity and due to the short training periods, we perform a simple 3-fold CV and rely on the highly efficient R-implementation of the elastic net "glmnet" of [Friedman et al., 2010]. In the conceptual introduction of the elastic net, we stressed the potential drawback of having no closed form solution. Consequently, we expect the model to perform slightly worse in small samples.

Regularized Synthetic Control (REGSC)

The fourth model under consideration is our proposed regularized synthetic control model which can be interpreted as a mixture of the elastic net and the original SC estimator. It is comparable to the elastic net insofar that it substitutes the Lasso-shrinkage by the inverse-Ridge-shrinkage. This substitution is motivated by the SC-specificity of having percent-like coefficients. Thus, we expect the model to perform well especially in settings that are comparable to the original SC setting like the static factor DGP. Its closed form solution and the increased flexibility caused by the tuneable hyperparameters make us confident that the model performs equally well in small and in large samples as well as in the pre- and the post-treatment period. To reduce computation time, the optimal hyperparameters are obtained by 2-fold CV in a two-step random grid-search procedure. Random hyperparameter grid search has proven to be more efficient than manual grid search both theoretically and empirically. See for instance [Bergstra and Bengio, 2012] for a careful discussion of hyperparameter optimization. Specifically, we start by spaning a large 50×50 grid with λ_1 ranging from 5 to 3,125 and λ_2 ranging from 10 to 10^7

and randomly select 400 out of the 2,500 λ_1-λ_2-combinations.[8] In the first step of the procedure, we identify the optimal hyperparameter combination of the initialized grid by applying 2-fold CV. Based on the result of the first step, we enclose the potential optimum in the second step by sequentially holding the first and the second hyperparameter fixed while increasing and decreasing the remaining hyperparameter on a coarser grid. Note that a more efficient CV-procedure could potentially improve the REGSC-method.

Factor model (FACTOR)

The data of this process is generated such that the common factors and the idiosyncratic component are uncorrelated and that the idiosyncratic errors are mutually uncorrelated. Thus, the static factor model which estimates the common factors as linear function of the donors is the most natural candidate model. The Principal Components (PC) estimator is a popular estimator for the factor model. It is employed as follows: In the pre-treatment period, we obtain the predictions by regressing the treatment series on the "latent" factors. As we implemented a two-factor structure, the factors are computed by multiplying the first two eigenvalues of the covariates matrix of the donors with the matrix of the covariates of the donors. The forecasts for the post-treatment period are obtained by multiplying the factor structure of the post-treatment period with the regression coefficients of the pre-treatment regression. As this model directly build upon the DGP, it is our benchmark-model and we expect it to perform best among all candidates.

For each of the 9 combinations of pre- and post-treatment period length and the 6 investigated donor group sizes, we simulated 1,000 static factor processes as described above.[9] This extensive simulation provides us with a total of 54,000 processes that are analyzed with respect to the following location and dispersion metrics in the post-treatment period: The Root Mean Squared Error (RMSE) is the central loss function of our analysis. It has the following form:

$$RMSFE(m) = \left(\frac{1}{T - T_0} \sum_{t=T_0}^{T} (y_{0,t} - \delta_{0,t} - \widehat{y}_{0,t}(m))^2 \right)^{1/2},$$

where m presents one of the five employed models. Due to its quadratic nature, it is not only a reasonable approximation to realistic loss structures but also mathematically convenient [Diebold, 2017]. Note that the treatment effect $\delta_{0,t}$ is subtracted from $y_{0,t}$ to re-calibrate the process towards its origin. The RMSE is unable to distinguish between over- and underestimation as deviations from the true target quantity are squared. The bias is directly related to the RMSE but provides a more detailed measure in terms of

[8] In sensitivity checks, we found that wider intervals did not change the location of the potential optimum. However, the optimal range will always be case-specific.

[9] Remember that $T_{pre} \in \{20, 50, 100\}$, $T_{post} \in \{10, 20, 30\}$ and $J \in \{5, 10, 15, 20, 25, 30\}$.

error location. It is computed as follows:

$$BIAS_m = \frac{1}{T - T_0} \sum_{t=T_0}^{T} \widehat{y}_m - (y_{0,t} - \delta_{0,t}),$$

such that negative values indicate under- and positive values overestimation. This precision metric is especially important when analyzing intercept-free models as these models will exhibits a bias whenever the treatment intercept falls outside the donor intercepts. The Mincer-Zarnowitz (MZ) regression tests the forecast optimality in a different and more holistic way by regressing the true value on its predicted value in the post-treatment period:

$$y_{0,t} = \beta_0 + \beta_1 \widehat{y}_{0,t} \text{ for } t \geq T - T_0.$$

If the forecast is optimal, we expect to observe that $(\beta_0, \beta_1) = (0, 1)$, an hypothesis that is directly testable by a simple F-test. Similar to the RMSE we can report average, simulation-overarching quantities without wiping out crucial details like negative and positive biases. In our simulation, we report the share of iterations in which the F-test accepted the joint hypothesis of $(\beta_0, \beta_1) = (0, 1)$ at the conventional significance level of 5%. The closer this share to unity, the closer the specific forecast to the optimal quantity. Due to varying sample sizes, however, those shares are only interpretable at the within-simulation level. To observe the variability of the employed models, we also compute their intra-simulation variances as

$$VAR(m) = \left(\frac{1}{T - T_0} \sum_{t=T_0}^{T} \widehat{y}_{0,t}(m) - \overline{\widehat{y}}_{0,t}(m) \right)^2.$$

4.1.3. Results

The full simulation results can be found in table format in the appendix 8.3.1 where we group the tables at the level of the six analyzed donor quantities. Here, we only present the central tendencies of the simulation. To do so, consider the following figure that plots the average RMSE of the models against the size of the donor pool for $T_{pre} = 50$ and $T_{post} \in \{10, 20, 30\}$.[10]

[10] As we found that the length of the post-treatment period is less important to explain the models performance, we pool the three elements of T_{post} at this point. The depicted means are thus computed on the basis of 3,000 iterations each. The same figures can be found in appendix 8.3.1 for $T_{pre} = 20$ and $T_{pre} = 100$.

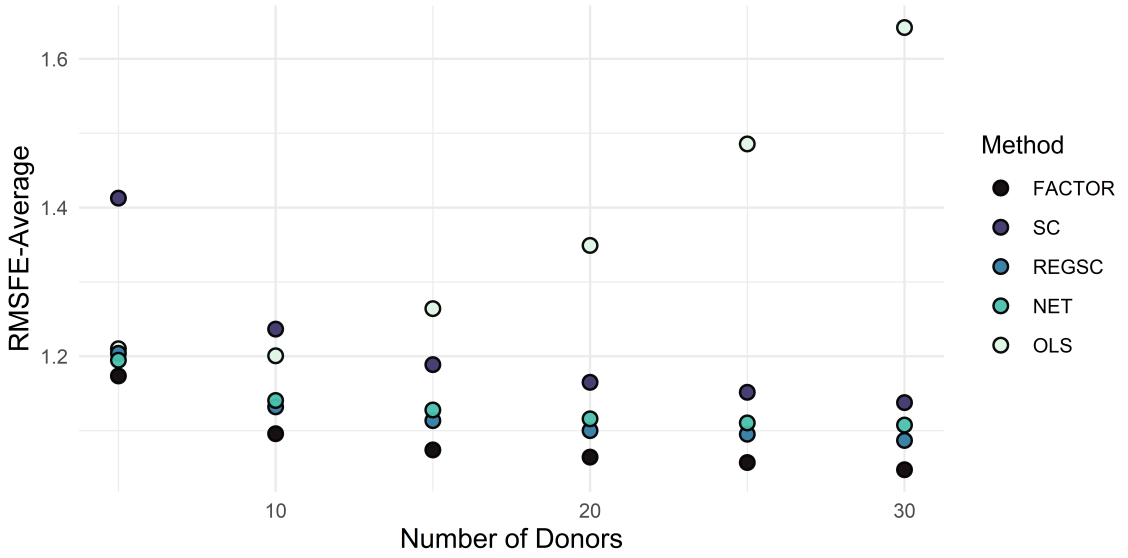

Figure 3. Simulation Performance for $T_{pre} = 50$ and $T_{post} \in \{10, 20, 30\}$

Five observations stand out: First, as expected, the unrestricted OLS regression starts to overfit the pre-treatment data quite fast indicated by a RMSE that starts to increase from $J = 10$ onward. This tendency consistently aggravates as J approaches T_{pre}. Second, the remaining models seem to successfully distinguish between systematic pre-treatment patterns and noise as their forecasts improve with increasing J. Third, in terms of relative performance, for obvious reasons, the factor models extrapolates the process best into the post-treatment period. Fourth, the elastic net and the proposed REGSC estimator perform comparable with slight comparative advantages toward the latter. These differences are more pronounced as J increases. Last but not least, the SC estimator, though being able to not overfit the training data, consistently performs worse than the elastic net and the REGSC estimator.

It has been stressed that the RMSE should not be the only precision metric when evaluating forecast performances as it is unable to detect over- and underestimation. On the other hand, the bias, though being able to flag iteration-specific over- and underestimations, can indicate spurious optimality when it is aggregated into a single metric like the mean.[11] To solve this problem, we first focus on the bias-distributions: Above, we argued, that intercept-free models like the SC-method will exhibit a positive (negative) bias whenever the mean of the treatment-series exceeds (falls below) the means of the donor series. In our simulation, the intercepts of the series are independent and identically distributed (iid) realizations of standard normal distributions. Therefore, the probability that the mean of the treatment series falls below (exceeds) all donor means equals $\frac{1}{J+1}$.[12]

[11] Consider for instance a model that forecasts $\{-1, 1\}$, each in 50% of the cases for a quantity whose optimal forecast is 0 in 100% of the cases. This model is far from optimal but the mean bias is 0.

[12] For each donor quantity J, there are $(J+1)!$ total orderings. In $\frac{(J+1)!}{J+1}$ of the cases, the intercept of the treatment series is the most extreme. This translate to a probability of $\frac{1}{J+1}$.

In the following figure, we extracted the iterations for which the treatment intercept was the most extreme (for all donor group sizes J), grouped these observations according to minimum (positive bias)/ maximum (negative bias) and plotted the post-treatment bias density for the SC and the REGSC model.

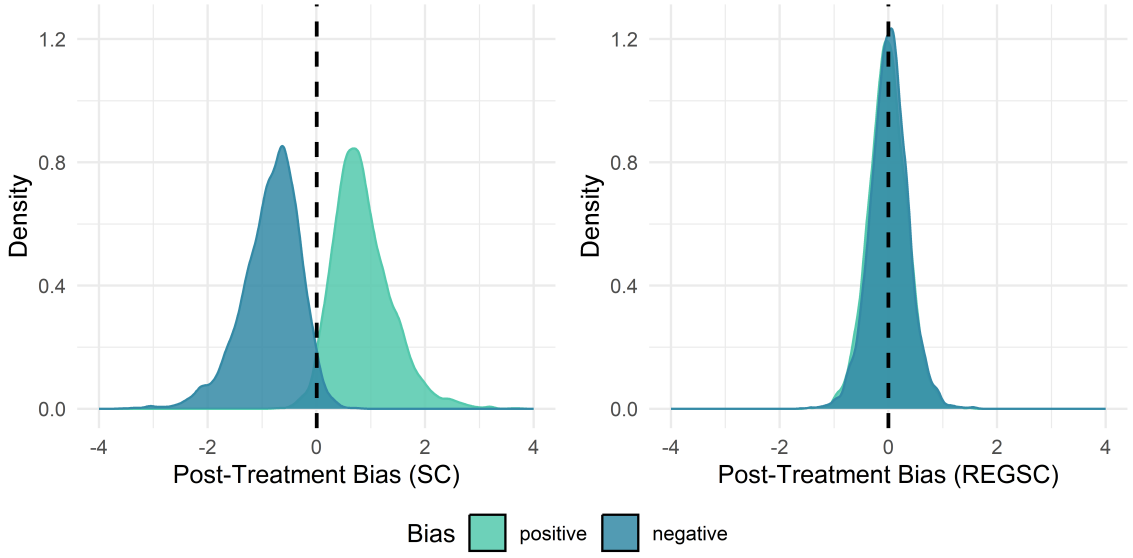

Figure 4. Bias-distribution for the SC and the REGSC model

The aforementioned bias-problem becomes immediately apparent: Though both estimators have an average bias that is close to 0, the SC method exhibits severe positive/ negative biases if the treatment intercept falls outside the convex hull of the donors intercepts. In contrast, the REGSC model as well as the remaining models that include an intercepts do not face this issue. In appendix 8.3.1, we group the observations at the model and the T_{pre} level and plot the corresponding densities. On a summarized level, we see, as expected, that the bias-distributions of all models become more tightly centered around zero the longer the prer-treatment periods and that the factor model, which serves as a benchmark model, outperforms all other models. Among these models the elastic net and the REGSC model are the closest contenders. This pattern also becomes visible when we consider the MZ-acceptance rate as in the following table:[13]

[13] As the benchmark factor model performs best in all cases, we report the second best model in column 8.

Table S1. MZ-Acceptance rates for $T_{pre} \in \{20, 50, 100\}$ and $J \in \{5, 10, 15, 20, 25, 30\}$

T_{pre}	Donors	SC	OLS	REGSC	NET	FACTOR	2nd best
20	5	0.5650	0.6617	0.7413	0.7648	0.8143	NET
20	10	0.6897	0.3777	0.7577	0.7479	0.8050	REGSC
20	15	0.7343	0.1153	0.7517	0.7486	0.8190	REGSC
20	20	0.7600	NA	0.7380	0.7193	0.7993	SC
20	25	NA	NA	0.7427	0.7420	0.7987	REGSC
20	30	NA	NA	0.7433	0.7475	0.8227	NET
50	5	0.5857	0.8707	0.8677	0.8820	0.8947	NET
50	10	0.7503	0.8237	0.8787	0.8783	0.8940	REGSC
50	15	0.8067	0.7123	0.8703	0.8680	0.8923	REGSC
50	20	0.8103	0.6007	0.8757	0.8717	0.8910	REGSC
50	25	0.8423	0.4333	0.8803	0.8750	0.8920	REGSC
50	30	0.8510	0.3023	0.8680	0.8667	0.8923	REGSC
100	5	0.6200	0.9100	0.9057	0.9150	0.9193	NET
100	10	0.7757	0.9120	0.9207	0.9277	0.9277	NET
100	15	0.8030	0.8743	0.9140	0.9067	0.9183	REGSC
100	20	0.8430	0.8480	0.9180	0.9157	0.9260	REGSC
100	25	0.8540	0.8097	0.9167	0.9117	0.9247	REGSC
100	30	0.8793	0.7537	0.9177	0.9127	0.9257	REGSC

For circumstances of few donors or cases where $T_{pre} \gg J$, the Lasso-penalty seems to be superior to the inverse-Ridge-penalty as in those cases, the elastic net outperforms the REGSC at a small margin. Further, we see that the SC regularization demonstrates its strength particularly in settings of many donors and few training observations but exhibits weaknesses in cases of few donors. For example, for $T_{pre} = 20$, the model improves its MZ-rate from 0.5650 to 0.7600 by about 135%. No other model features such high absolute and relative improvements. If $\frac{T_{pre}}{J} < 1$, the simple SC approach without covariates struggles with parameter identification. Consequently, it's not possible to compute MZ-rates for rows 5 and 6. Note, however, that the original SC-method does not face such limitations and is able to compute the synthetic control even when $J \gg T_{pre}$.

The static simulation study can be concluded as follows: With exception of the OLS model, all employed models are capable to distinguish systematic pre-treatment patterns from noise and do not overfit the training data. The simplified version of the SC method is particularly promising, when $\frac{T_{pre}}{J} \approx 1$, however, as stressed by [Abadie et al., 2010] it is of crucial importance to verify that the treatment intercept lies inside the bandwidth of the donor intercepts as substantial biases results if ignored. The elastic net and REGSC model include an intercept and allow more flexible weight coefficients and can thus be applied independently of the mean characteristics of the series. In particular, the former promises high forecast accuracy when the donor pool is small. In the remaining 12 out of 18 considered cases, we find that our proposed REGSC estimator performs best. A similar conclusion can also be drawn on the basis of the RMSE instead of the MZ-acceptance rates.

The corresponding is postponed to appendix 8.3.1. Before we test our model outside the world of synthesized data, we show the simulation results of a dynamic DGP that is inspired by the real world covariance structure of the GDPs of the G20 countries.

4.2. Dynamic Data Generating Processes

5. Simulation

5.1. Static Data Generating Processes

Simulation-Procedure

5.2. Weakly Dynamic Data Generating Processes

5.3. Dynamic Data Generating Processes

In order to rigorously evaluate the performance of both our proposed and existing SC estimators, a key milestone would be to test the estimators in simulated datasets which mimic the real world. As many of the previous studies have focused on the economic development following a treatment, it stands to reason that real world changes in GDP could serve as the basis for a close-to-reality inspired DGP. To ensure that the DGP is based on a relatively uniform reference group which inhibits significant amounts of commonalities and correlation, the data basis consists of all countries from the G20 as well as the European Union (EU). [14] Achtung: R Paket richtig zitieren
The dataset is subjected to two additional filters: firstly, only countries with at least 40 years worth of GDP data remain in the dataset, and secondly, the time series selected must be stationary. The latter is tested using the Augmented Dickey Fuller Test on a 10% significance level. The remaining 22 countries are the base from which the close-to-reality datasets are simulated using a VAR model.
Two different approaches, which will be referred to in the following as the *micro approach* and *macro approach*, are being considered. To evaluate the proposed models in terms of an increasing number of donors as in the static case, the results naturaly vary with the the underlying base, the data is simulated from. It depends critically on whether the simulation is perfomred with the equal number of donors as the models are estimated with (micro approach), or whether the simulation is performed with a larger number of donors such that the model estimation is carried out with subsets of the whole simulation (macro approach). The micro approach is therefore a more controlled and stylized design, in which the model evaluation is performed on the full simulated dataset, i.e. is therefore internally consistent. In this case, adding more donors leads to an increasingly complex model. The macro approach is a more realistic approach in which the effects of adding new information to the model estimation in terms of donors become apparent while the simulated dataset remains structurally the same.

, depending on the underlying base, the data is simulated from. To evaluate the proposed models in terms of increasing number of Donors, as in the static case, it In the *macro* view, the data is si

Micro View

[14] The GDP data is sourced from the World Bank's World Development Indicators, which is directly accessible through the WDI-Package in R using the ticker 'NY.GDP.PCAP.KD.ZG' (GDP Per Capita Growth Rate).

Macro View

6. Applications

to verify: elastic net needs to perform time series CV. Not relevant for REGSC as df is split from 1 onward. Also relevant for VAR simulations

We consider three leading examples:

6.1. The Economic Costs of Conflict

[Abadie and Gardeazabal, 2003]

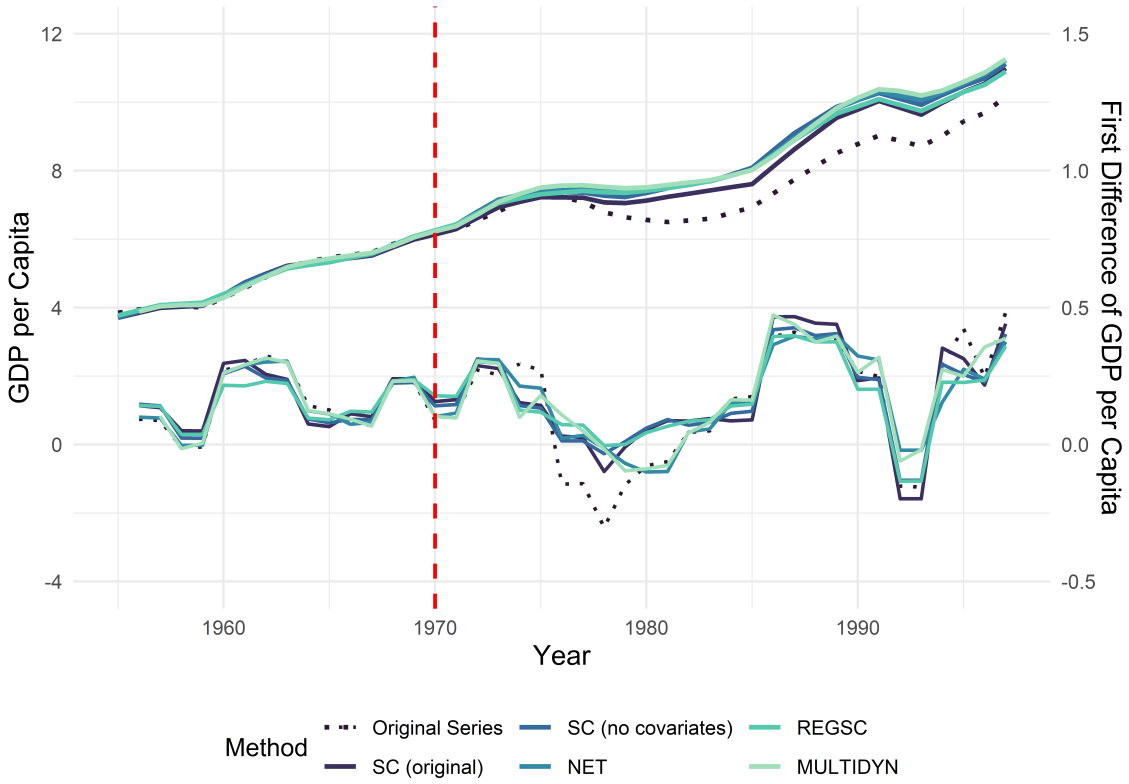

Figure 5. GDP per Capita for the (synthetic) Basque

6.2. Estimating the Effect of California's Tobacco Control Program

[Abadie et al., 2010]

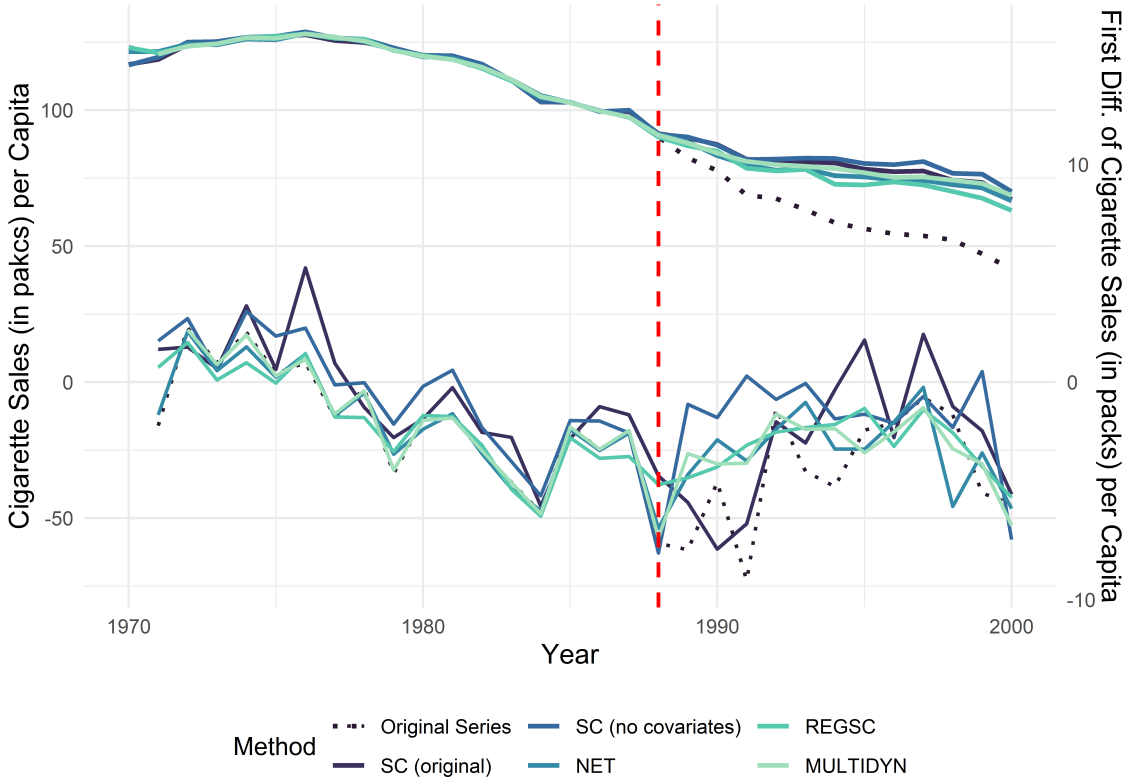

Figure 6. Cigarette Sales per Capita for the (synthetic) California

6.3. The Economic Cost of the 1990 German Reunification

[Abadie et al., 2015]

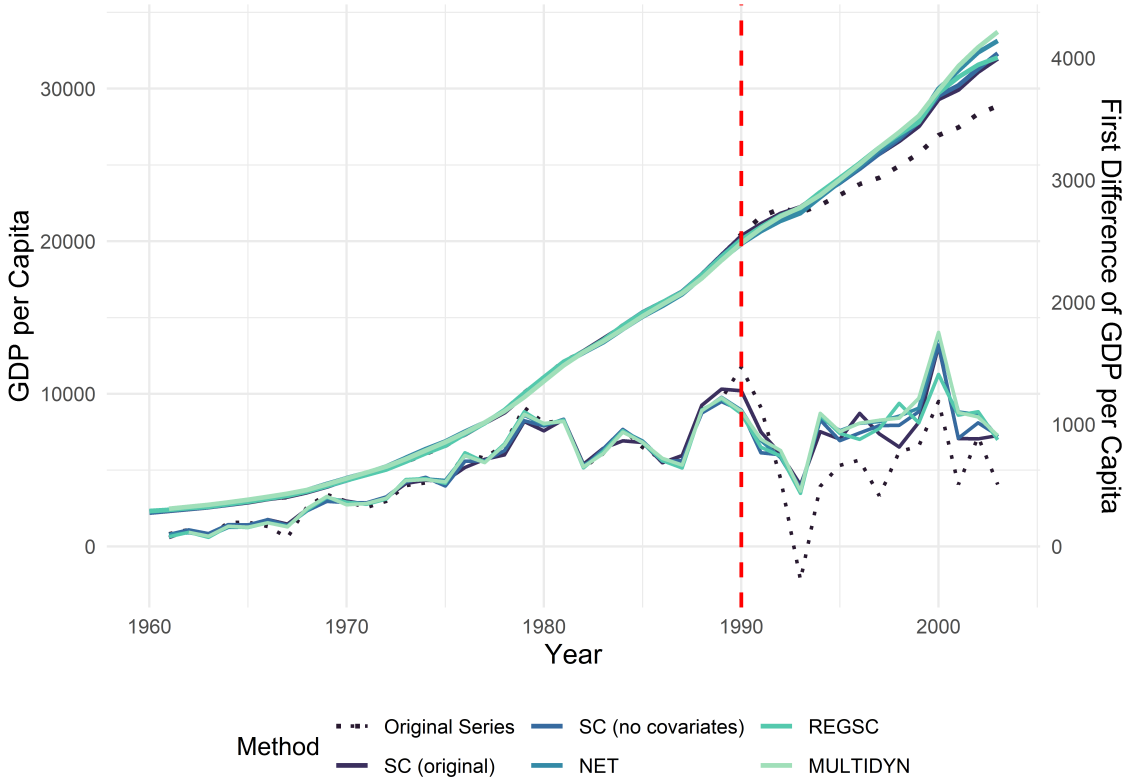

Figure 7. GDP per Capita for the (synthetic) West Germany

7. Conclusion

- Some concluding remarks and an outlook
- Keep short, around 1-2 pages
- Natural extension: case with explanatory variables
- We advocate for an interval instead of an point forecast. Therefore also for an interval estimate of the treatment effect. Report more measures of uncertainty than permutation p-values.

References

[Abadie et al., 2010] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105:493–505.

[Abadie et al., 2015] Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

[Abadie and Gardeazabal, 2003] Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132.

[Abadie and L'Hour, 2021] Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116:1–34.

[Andrews, 2003] Andrews, D. W. K. (2003). End-of-Sample Instability Tests. *Econometrica*, 71(6):1661–1694.

[Athey and Imbens, 2016] Athey, S. and Imbens, G. (2016). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives*, 31.

[Bekker, 1994] Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–81.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *SSRN Electronic Journal*.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84.

[Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13(null):281–305.

[Born et al., 2019] Born, B., Müller, G. J., Schularick, M., and Sedláček, P. (2019). The Costs of Economic Nationalism: Evidence from the Brexit Experiment*. *The Economic Journal*, 129(623):2722–2744.

[Breitung and Knüppel, 2021] Breitung, J. and Knüppel, M. (2021). How far can we forecast? Statistical tests of the predictive content. *Journal of Applied Econometrics*, 36(4):369–392.

[Chernozhukov et al., 2021] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls.

University of California at San Diego, Economics Working Paper Series qt90m9d66s, Department of Economics, UC San Diego.

[Cho, 2020] Cho, S.-W. S. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3):323–344.

[Chow, 1960] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

[Cunningham, 2021] Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

[Diebold, 2017] Diebold, F. X. (2017). *Forecasting*. Department of Economics, University of Pennsylvania.

[Doudchenko and Imbens, 2016] Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc.

[Ferman, 2021] Ferman, B. (2021). On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

[Fisher, 1935] Fisher, R. A. (1971 [1935]). The design of experiments (9th ed.). *Macmillian*.

[Friedman et al., 2010] Friedman, J. H., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(i01).

[Harvey and Thiele, 2020] Harvey, A. and Thiele, S. (2020). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36.

[Hoerl and Kennard, 1970] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

[Kirchgässner et al., 2012] Kirchgässner, G., Wolters, J., and Hassler, U. (2012). *Introduction to Modern Time Series Analysis*. Springer Texts in Business and Economics.

[Koop, 2003] Koop, G. (2003). *Bayesian Econometrics*. Wiley.

[Muhlbach and Nielsen, 2019] Muhlbach, N. S. and Nielsen, M. S. (2019). Tree-based Synthetic Control Methods: Consequences of moving the US Embassy. Papers 1909.03968, arXiv.org.

[Neyman, 1923] Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*, 4:465–472.

[Rubin, 1974] Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288.

8. Appendix

8.1. Simple Static Extension

8.1.1. OLS Solution

In case the population covariance matrix is observable, the OLS-coefficients can be directly derived from it: $\left(w_1^{OLS}, w_2^{OLS}\right) = \mathbf{\Sigma_2}^{-1}\boldsymbol{\sigma_{12}}$

8.1.2. SC Solution

The restricted solution is can directly be derived from the covariance matrix. The first index in the square brackets indicates the row, the second the column position.

$$\begin{aligned} w_1^{SC} &= (\boldsymbol{\sigma_{12}}'[1] - \boldsymbol{\sigma_{12}}'[2] - \mathbf{\Sigma_2}[2,1] + \mathbf{\Sigma_2}[1,1]) \,/\, (\mathbf{\Sigma_2}[1,1] + \mathbf{\Sigma_2}[2,2] - 2 * \mathbf{\Sigma_2}[1,2]) \\ &= (0.1 - 0.4 - 0.5 + 1)/(1 + 1 - 2 * 0.5) = 0.2 \end{aligned}$$

$$\begin{aligned} w_2^{SC} &= (\boldsymbol{\sigma_{12}}'[2] - \boldsymbol{\sigma_{12}}'[1] - \mathbf{\Sigma_2}[1,2] + \mathbf{\Sigma_2}[2,2]) \,/\, (\mathbf{\Sigma_2}[2,2] + \mathbf{\Sigma_2}[1,1] - 2 * \mathbf{\Sigma_2}[2,1]) \\ &= (0.4 - 0.1 - 0.5 + 1)/(1 + 1 - 2 * 0.5) = 0.8 \end{aligned}$$

8.1.3. Variances

The variances are derived from the weights and the covariance matrix:

$$\begin{aligned} var(Y_0 - w_1 Y_1 - w_2 Y_2) = var(Y_0) + w_1^2 \cdot var(Y_1) + w_2^2 \cdot var(Y_2) \; - \\ 2 \cdot w_1 \cdot cov(Y_0, Y_1) - 2 \cdot w_2 \cdot cov(Y_0, Y_2) \; + \\ 2 \cdot w_1 w_2 \cdot cov(Y_1, Y_2) \end{aligned}$$

8.2. General Static Extension

8.2.1. REGSC: The limit for $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$

For $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$ the objective function reduce to

$$Q(\lambda_1, \lambda 2) = \lambda_1 w'w + \lambda_2 (1 - \mathbf{1}'w)^2$$

The derivative is obtained as

$$\frac{\partial Q(\lambda_1, \lambda_2)}{\partial w} = 2\lambda_1 w + 2\lambda_2(\mathbf{1} - \mathbf{1}\mathbf{1}'w)$$

By setting the derivative to zero and multiplying with $\mathbf{1}$ we obtain:

$$\lambda_1 \mathbf{1}'w + \lambda_2(n - \mathbf{1}'w) = 0$$

where $\mathbf{1}'w = \sum w_i$. Solving for $\mathbf{1}'w$ we obtain

$$\mathbf{1}'w = \frac{1}{1 + \lambda_1/\lambda_2}$$

and due to the symmetry of the objective function with respect to the elements of the weight vector we have

$$w_i = 1/(n + n\lambda_1/\lambda_2)$$

8.3. Simulation Study

8.3.1. Static Simulation results

Table S2. Simulation Results of the Static Factor Model with **J = 5** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.2497 (-0.0045) [0.7250] {0.6523}	1.4455 (-0.0086) [0.4210] {1.1179}	1.3064 (-0.0144) [0.6240] {0.6898}	1.3086 (-0.0073) [0.6691] {0.6581}	1.3814 (-0.0098) [0.5140] {1.1678}
	20	1.2509 (0.0137) [0.8240] {0.6143}	1.4575 (0.0331) [0.5360] {1.0690}	1.3054 (0.0063) [0.7460] {0.6619}	1.3048 (0.0085) [0.7516] {0.6137}	1.3661 (0.0114) [0.6650] {1.1028}
	10	1.2344 (-0.0092) [0.8940] {0.6137}	1.4283 (-0.0120) [0.7380] {1.0366}	1.2847 (-0.0075) [0.8540] {0.6505}	1.2804 (-0.0098) [0.8723] {0.6086}	1.3520 (-0.0126) [0.8060] {1.0948}
50	30	1.1831 (0.0147) [0.8510] {0.6459}	1.4185 (-0.0223) [0.4600] {1.0600}	1.2131 (0.0137) [0.8210] {0.6440}	1.2045 (0.0117) [0.8388] {0.5834}	1.2190 (0.0106) [0.8250] {0.7974}
	20	1.1740 (-0.0195) [0.9030] {0.6426}	1.4160 (-0.0106) [0.5630] {1.0680}	1.2044 (-0.0204) [0.8650] {0.6311}	1.1948 (-0.0191) [0.8810] {0.5752}	1.2105 (-0.0166) [0.8750] {0.7879}
	10	1.1635 (-0.0081) [0.9300] {0.5963}	1.4033 (-0.0211) [0.7340] {1.0002}	1.1940 (-0.0115) [0.9170] {0.5850}	1.1845 (-0.0081) [0.9260] {0.5407}	1.2009 (-0.0074) [0.9120] {0.7415}
100	30	1.1645 (0.0047) [0.8920] {0.6494}	1.3955 (0.0381) [0.4950] {1.0668}	1.1811 (0.0053) [0.8700] {0.6405}	1.1733 (0.0048) [0.8930] {0.5931}	1.1799 (0.0052) [0.8870] {0.7238}
	20	1.1576 (0.0026) [0.9220] {0.6389}	1.3796 (-0.0384) [0.6080] {1.0337}	1.1756 (0.0040) [0.9100] {0.6226}	1.1691 (0.0021) [0.9120] {0.5871}	1.1755 (0.0012) [0.9100] {0.7090}
	10	1.1518 (0.0108) [0.9440] {0.5971}	1.3907 (0.0592) [0.7570] {0.9826}	1.1682 (0.0156) [0.9370] {0.5740}	1.1643 (0.0146) [0.9400] {0.5432}	1.1697 (0.0135) [0.9330] {0.6623}

Table S3. Simulation Results of the Static Factor Model with **J = 10** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.1575 (-0.0071) [0.7090] {0.7830}	1.2944 (0.0002) [0.5720] {1.0468}	1.2229 (-0.0047) [0.6590] {0.7805}	1.2521 (-0.0093) [0.6416] {0.7714}	1.6286 (0.0018) [0.1790] {2.3061}
	20	1.1611 (0.0060) [0.8220] {0.7459}	1.2759 (0.0197) [0.6920] {1.0523}	1.2336 (0.0071) [0.7590] {0.7777}	1.2591 (0.0036) [0.7495] {0.7613}	1.6075 (0.0045) [0.3520] {2.1785}
	10	1.1675 (0.0040) [0.8840] {0.6883}	1.2963 (0.0127) [0.8050] {0.9859}	1.2318 (0.0004) [0.8550] {0.7180}	1.2659 (0.0006) [0.8513] {0.6984}	1.6276 (0.0181) [0.6020] {2.0705}
50	30	1.1048 (0.0049) [0.8640] {0.7755}	1.2511 (0.0020) [0.6680] {1.0148}	1.1444 (0.0079) [0.8240] {0.7207}	1.1524 (0.0065) [0.8310] {0.6795}	1.2140 (0.0087) [0.7420] {1.1008}
	20	1.0985 (0.0047) [0.8870] {0.7744}	1.2416 (0.0037) [0.7270] {0.9962}	1.1327 (0.0029) [0.8750] {0.7333}	1.1417 (0.0042) [0.8750] {0.6573}	1.2022 (0.0028) [0.8190] {1.0658}
	10	1.0838 (0.0001) [0.9310] {0.7704}	1.2171 (0.0033) [0.8560] {0.9989}	1.1183 (-0.0039) [0.9370] {0.7199}	1.1270 (-0.0034) [0.9290] {0.6649}	1.1857 (-0.0049) [0.9100] {1.0604}
100	30	1.0920 (0.0040) [0.9070] {0.8025}	1.2323 (0.0119) [0.6930] {1.0112}	1.1141 (0.0050) [0.9040] {0.7620}	1.1171 (0.0039) [0.9080] {0.6968}	1.1405 (0.0025) [0.8780] {0.9363}
	20	1.0842 (-0.0057) [0.9360] {0.7858}	1.2232 (-0.0031) [0.7680] {1.0095}	1.1073 (-0.0049) [0.9240] {0.7427}	1.1088 (-0.0064) [0.9280] {0.6902}	1.1333 (-0.0081) [0.9150] {0.9314}
	10	1.0742 (0.0080) [0.9400] {0.7377}	1.2059 (-0.0206) [0.8660] {0.9476}	1.0919 (0.0021) [0.9340] {0.6925}	1.0986 (0.0016) [0.9470] {0.6526}	1.1186 (0.0006) [0.9430] {0.8653}

Table S4. Simulation Results of the Static Factor Model with **J = 15** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.1398 (0.0243) [0.7340] {0.7617}	1.2604 (0.0374) [0.6240] {1.0341}	1.2276 (0.0244) [0.6450] {0.8403}	1.2533 (0.0197) [0.6408] {0.7997}	2.4777 (0.0312) [0.0320] {6.3624}
	20	1.1287 (0.0042) [0.8200] {0.7481}	1.2417 (0.0258) [0.7250] {0.9966}	1.2149 (0.0129) [0.7460] {0.8261}	1.2432 (0.0165) [0.7386] {0.7667}	2.4716 (-0.0097) [0.0690] {6.3819}
	10	1.1112 (0.0206) [0.9030] {0.7130}	1.2309 (-0.0240) [0.8540] {0.9484}	1.1973 (0.0224) [0.8640] {0.7702}	1.2167 (0.0247) [0.8660] {0.7234}	2.4393 (0.0245) [0.2450] {6.0769}
50	30	1.0818 (-0.0032) [0.8580] {0.8282}	1.1935 (0.0006) [0.7490] {1.0035}	1.1198 (-0.0019) [0.8230] {0.7778}	1.1356 (-0.0019) [0.8070] {0.7102}	1.2747 (-0.0040) [0.5980] {1.3448}
	20	1.0879 (-0.0133) [0.8950] {0.7705}	1.1975 (-0.0183) [0.8060] {0.9465}	1.1307 (-0.0205) [0.8710] {0.7340}	1.1470 (-0.0188) [0.8760] {0.6722}	1.2892 (-0.0210) [0.6930] {1.2960}
	10	1.0522 (-0.0132) [0.9240] {0.7913}	1.1754 (-0.0098) [0.8650] {0.9519}	1.0896 (-0.0130) [0.9170] {0.7423}	1.1004 (-0.0122) [0.9210] {0.6776}	1.2281 (-0.0011) [0.8460] {1.2605}
100	30	1.0678 (0.0007) [0.8950] {0.8395}	1.1818 (-0.0212) [0.7310] {0.9981}	1.0890 (0.0007) [0.8900] {0.7767}	1.0974 (-0.0004) [0.8730] {0.7073}	1.1415 (-0.0034) [0.8310] {1.0481}
	20	1.0667 (0.0048) [0.9290] {0.8321}	1.1809 (-0.0018) [0.8040] {0.9770}	1.0886 (0.0073) [0.9230] {0.7671}	1.0947 (0.0064) [0.9180] {0.6998}	1.1468 (0.0066) [0.8800] {1.0384}
	10	1.0411 (-0.0059) [0.9310] {0.7897}	1.1459 (-0.0049) [0.8740] {0.9161}	1.0654 (-0.0045) [0.9290] {0.7422}	1.0717 (-0.0057) [0.9290] {0.6802}	1.1200 (-0.0056) [0.9120] {0.9948}

Table S5. Simulation Results of the Static Factor Model with **J = 20** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.1228 (0.0074) [0.7090] {0.8229}	1.2229 (-0.0229) [0.6550] {1.0024}	1.2040 (0.0044) [0.6410] {0.8544}	1.2331 (-0.0010) [0.6079] {0.8238}	NA (NA) [NA] {NA}
	20	1.1230 (0.0009) [0.8030] {0.7731}	1.2276 (-0.0067) [0.7520] {0.9833}	1.2093 (-0.0008) [0.7250] {0.8563}	1.2385 (-0.0041) [0.7144] {0.8207}	NA (NA) [NA] {NA}
	10	1.0939 (0.0023) [0.8860] {0.7695}	1.1927 (-0.0226) [0.8730] {0.9365}	1.1749 (0.0033) [0.8480] {0.8217}	1.2036 (0.0031) [0.8357] {0.7697}	NA (NA) [NA] {NA}
50	30	1.0662 (-0.0121) [0.8640] {0.8580}	1.1629 (-0.0148) [0.7290] {0.9743}	1.1007 (-0.0141) [0.8330] {0.7912}	1.1180 (-0.0182) [0.8230] {0.7118}	1.3564 (-0.0184) [0.4080] {1.6096}
	20	1.0702 (0.0004) [0.8810] {0.8213}	1.1693 (0.0055) [0.8070] {0.9484}	1.1085 (-0.0004) [0.8710] {0.7632}	1.1205 (0.0008) [0.8690] {0.7049}	1.3538 (-0.0073) [0.5880] {1.5678}
	10	1.0566 (0.0204) [0.9280] {0.7780}	1.1627 (0.0143) [0.8950] {0.9080}	1.0907 (0.0234) [0.9230] {0.7208}	1.1086 (0.0242) [0.9230] {0.6595}	1.3371 (0.0204) [0.8060] {1.4765}
100	30	1.0512 (-0.0083) [0.9040] {0.8556}	1.1442 (0.0027) [0.7800] {0.9709}	1.0735 (-0.0072) [0.8820] {0.7933}	1.0856 (-0.0073) [0.8890] {0.7236}	1.1639 (-0.0066) [0.7880] {1.1355}
	20	1.0438 (-0.0035) [0.9230] {0.8657}	1.1434 (0.0063) [0.8380] {0.9643}	1.0652 (-0.0009) [0.9260] {0.7987}	1.0763 (-0.0030) [0.9200] {0.7379}	1.1584 (-0.0035) [0.8400] {1.1555}
	10	1.0381 (-0.0014) [0.9510] {0.8044}	1.1236 (-0.0054) [0.9110] {0.8937}	1.0615 (0.0005) [0.9460] {0.7387}	1.0729 (-0.0024) [0.9380] {0.6780}	1.1502 (-0.0035) [0.9160] {1.0692}

Table S6. Simulation Results of the Static Factor Model with **J = 25** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.1154 (0.0019) [0.6980] {0.8012}	NA (NA) [NA] {NA}	1.2097 (-0.0002) [0.6280] {0.8593}	1.2225 (-0.0036) [0.6440] {0.7918}	NA (NA) [NA] {NA}
	20	1.1038 (0.0081) [0.8150] {0.7882}	NA (NA) [NA] {NA}	1.2001 (0.0127) [0.7410] {0.8314}	1.2240 (0.0125) [0.7310] {0.7967}	NA (NA) [NA] {NA}
	10	1.1051 (0.0056) [0.8830] {0.7503}	NA (NA) [NA] {NA}	1.1829 (-0.0046) [0.8590] {0.7897}	1.2061 (-0.0038) [0.8513] {0.7621}	NA (NA) [NA] {NA}
50	30	1.0657 (0.0104) [0.8510] {0.8576}	1.1525 (-0.0014) [0.7880] {0.9696}	1.1039 (0.0121) [0.8390] {0.8090}	1.1198 (0.0119) [0.8270] {0.7223}	1.4937 (0.0138) [0.2260] {2.0239}
	20	1.0541 (0.0175) [0.9000] {0.8634}	1.1514 (0.0171) [0.8460] {0.9776}	1.0905 (0.0202) [0.8840] {0.8007}	1.1090 (0.0204) [0.8720] {0.7497}	1.4816 (0.0229) [0.3860] {2.0191}
	10	1.0513 (-0.0003) [0.9250] {0.8091}	1.1507 (0.0100) [0.8930] {0.9253}	1.0911 (-0.0041) [0.9180] {0.7763}	1.1025 (-0.0056) [0.9260] {0.7072}	1.4816 (-0.0103) [0.6880] {1.9764}
100	30	1.0377 (-0.0036) [0.9190] {0.8703}	1.1262 (0.0009) [0.8120] {0.9562}	1.0607 (-0.0028) [0.9070] {0.7995}	1.0719 (-0.0026) [0.9000] {0.7252}	1.1913 (0.0002) [0.7200] {1.2385}
	20	1.0431 (-0.0088) [0.9220] {0.8747}	1.1258 (-0.0108) [0.8450] {0.9594}	1.0656 (-0.0075) [0.9180] {0.8108}	1.0778 (-0.0084) [0.9110] {0.7317}	1.1897 (-0.0028) [0.8160] {1.2371}
	10	1.0240 (-0.0011) [0.9330] {0.8335}	1.1119 (0.0133) [0.9050] {0.9086}	1.0478 (0.0015) [0.9250] {0.7689}	1.0588 (0.0023) [0.9240] {0.6957}	1.1803 (0.0046) [0.8930] {1.1779}

Table S7. Simulation Results of the Static Factor Model with **J = 30** Donors.

T_{pre}	T_{post}	**FACTOR**	**SC**	**REGSC**	**NET**	**OLS**
		RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}	RMSE (BIAS) [MZ-RATES] {VARIANCE}
20	30	1.0985 (0.0061) [0.7480] {0.8283}	NA (NA) [NA] {NA}	1.1867 (0.0131) [0.6260] {0.8679}	1.2139 (0.0104) [0.6397] {0.8722}	NA (NA) [NA] {NA}
	20	1.0915 (0.0044) [0.8140] {0.8092}	NA (NA) [NA] {NA}	1.1714 (-0.0013) [0.7480] {0.8365}	1.2129 (0.0022) [0.7409] {0.8370}	NA (NA) [NA] {NA}
	10	1.0803 (0.0225) [0.9060] {0.7730}	NA (NA) [NA] {NA}	1.1591 (0.0147) [0.8560] {0.8143}	1.1885 (0.0210) [0.8616] {0.7921}	NA (NA) [NA] {NA}
50	30	1.0552 (0.0048) [0.8600] {0.8536}	1.1492 (0.0185) [0.8060] {0.9671}	1.0989 (0.0050) [0.8220] {0.7988}	1.1206 (0.0012) [0.8270] {0.7556}	1.6600 (0.0062) [0.0960] {2.5759}
	20	1.0488 (0.0021) [0.8990] {0.8787}	1.1312 (0.0090) [0.8560] {0.9519}	1.0819 (-0.0005) [0.8680] {0.8199}	1.1032 (-0.0001) [0.8630] {0.7509}	1.6398 (-0.0045) [0.2330] {2.5723}
	10	1.0381 (0.0031) [0.9180] {0.8092}	1.1327 (-0.0015) [0.8910] {0.8922}	1.0792 (-0.0019) [0.9140] {0.7594}	1.0985 (-0.0064) [0.9100] {0.6948}	1.6269 (-0.0201) [0.5780] {2.3713}
100	30	1.0460 (0.0116) [0.9060] {0.8951}	1.1253 (0.0154) [0.8500] {0.9630}	1.0709 (0.0107) [0.9020] {0.8098}	1.0840 (0.0118) [0.8890] {0.7518}	1.2419 (0.0059) [0.6290] {1.3709}
	20	1.0278 (-0.0039) [0.9380] {0.9037}	1.1067 (-0.0003) [0.8750] {0.9668}	1.0477 (-0.0035) [0.9270] {0.8308}	1.0626 (-0.0021) [0.9210] {0.7497}	1.2180 (0.0002) [0.7660] {1.3736}
	10	1.0300 (-0.0309) [0.9330] {0.8168}	1.0995 (-0.0335) [0.9130] {0.8818}	1.0509 (-0.0339) [0.9240] {0.7504}	1.0646 (-0.0330) [0.9280] {0.6825}	1.2299 (-0.0419) [0.8660] {1.2708}

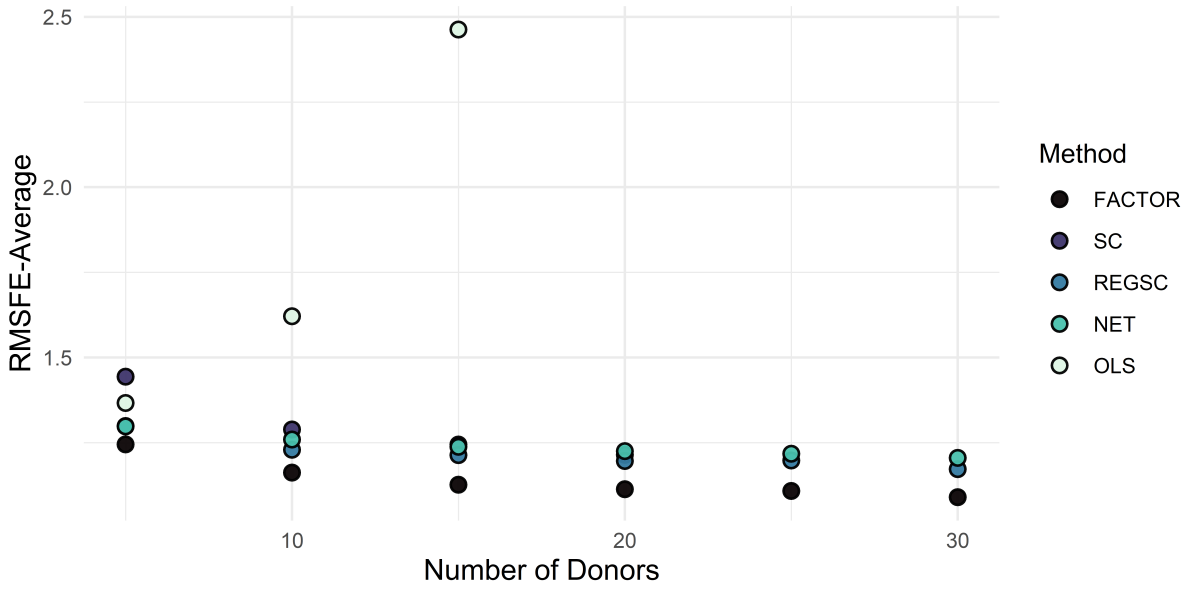

Figure 8. Simulation Performance for $T_{pre} = 20$ and $T_{post} \in \{10, 20, 30\}$

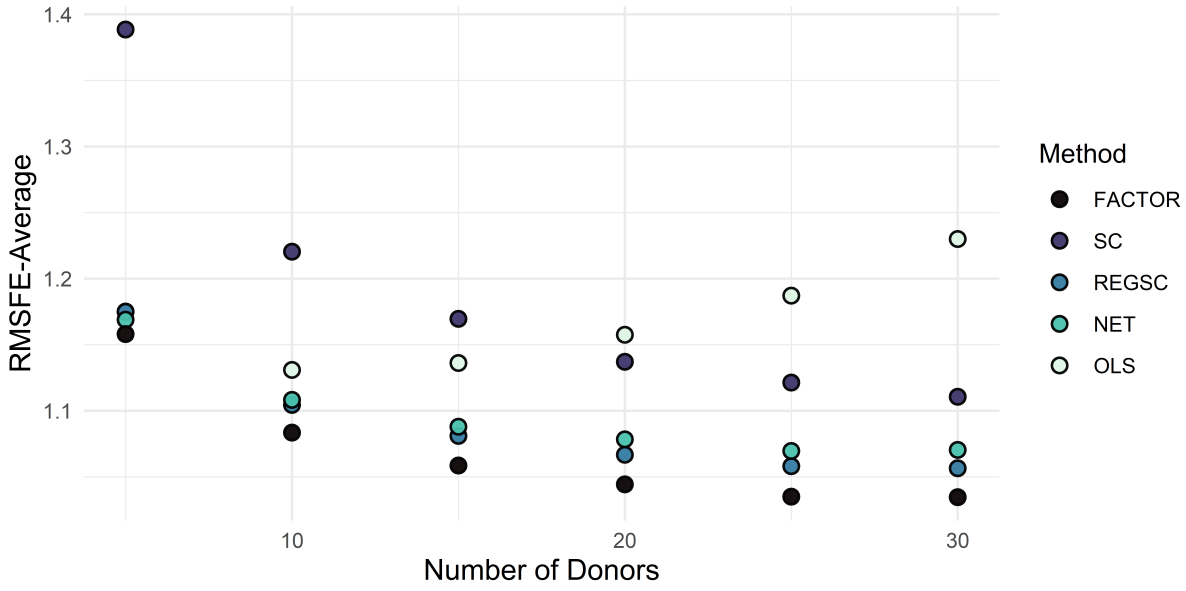

Figure 9. Simulation Performance for $T_{pre} = 100$ and $T_{post} \in \{10, 20, 30\}$

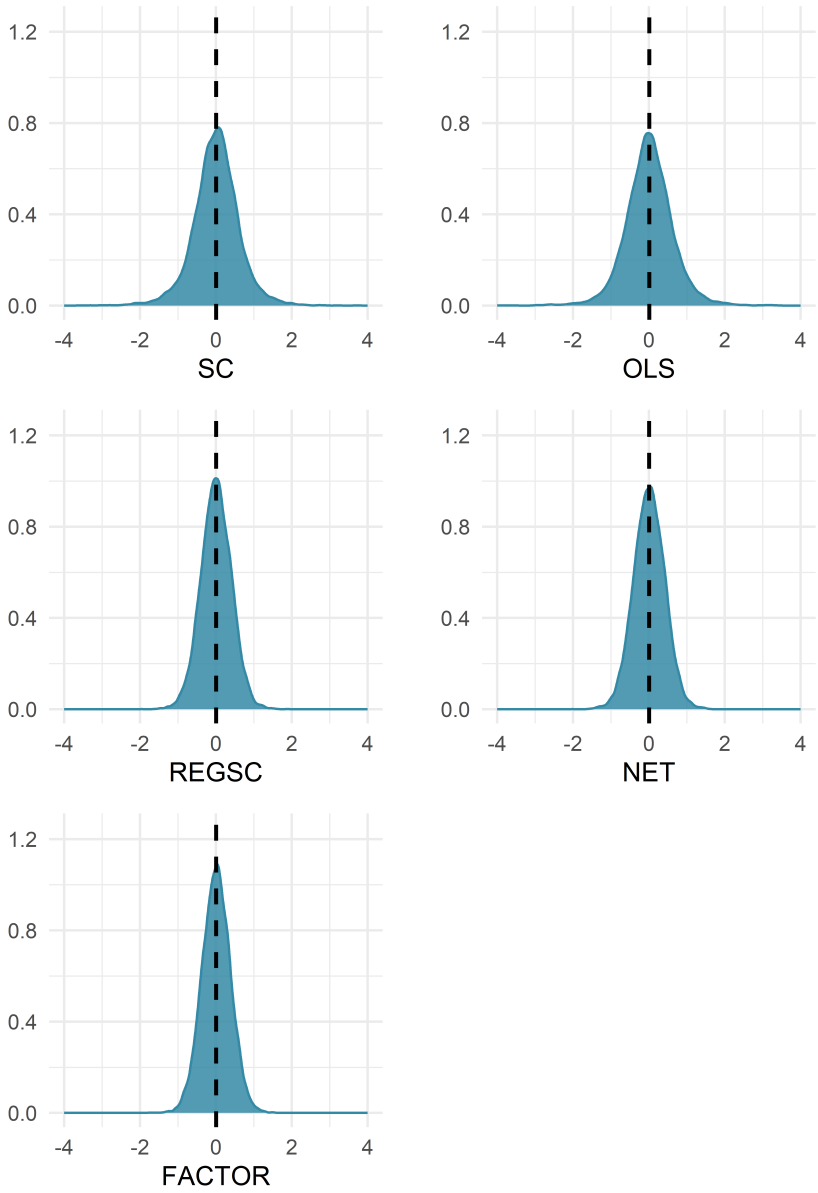

Figure 10. Bias-densities for $T_{pre} = 20$ and $T_{post} \in \{10, 20, 30\}$

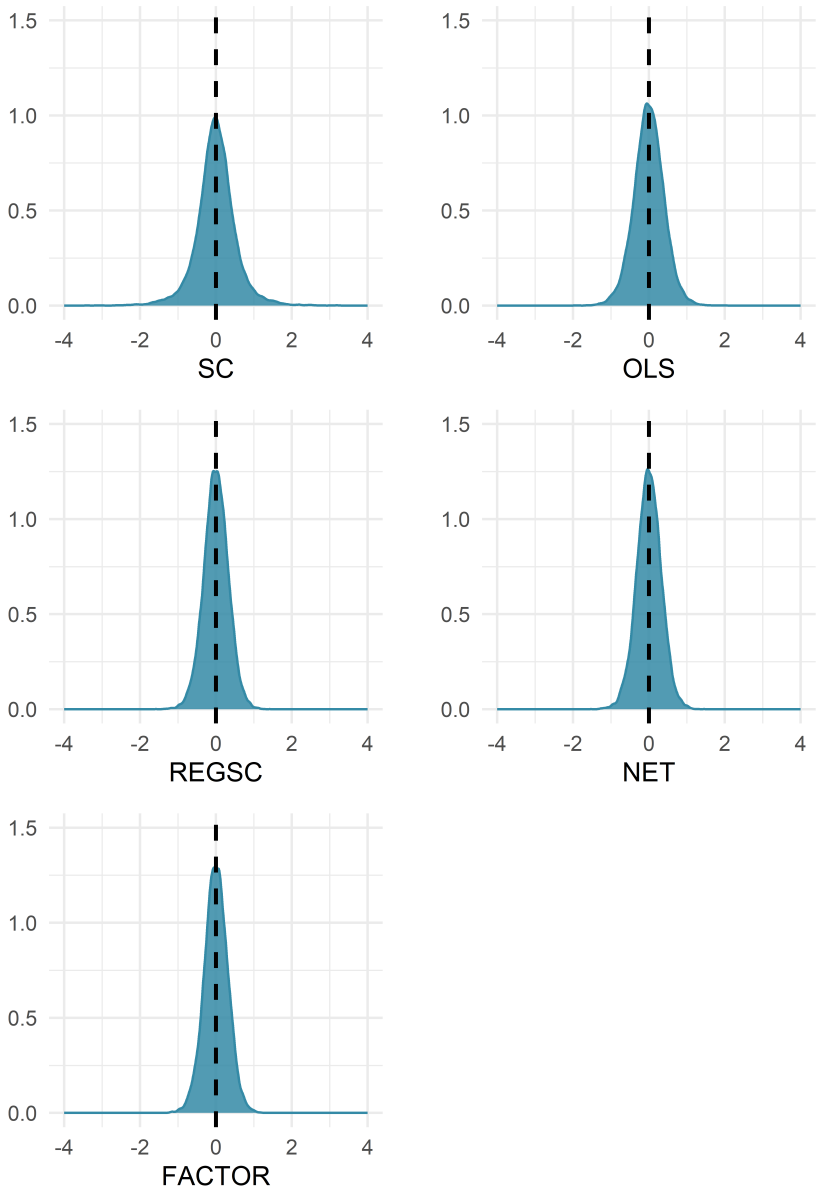

Figure 11. Bias-densities for $T_{pre} = 50$ and $T_{post} \in \{10, 20, 30\}$

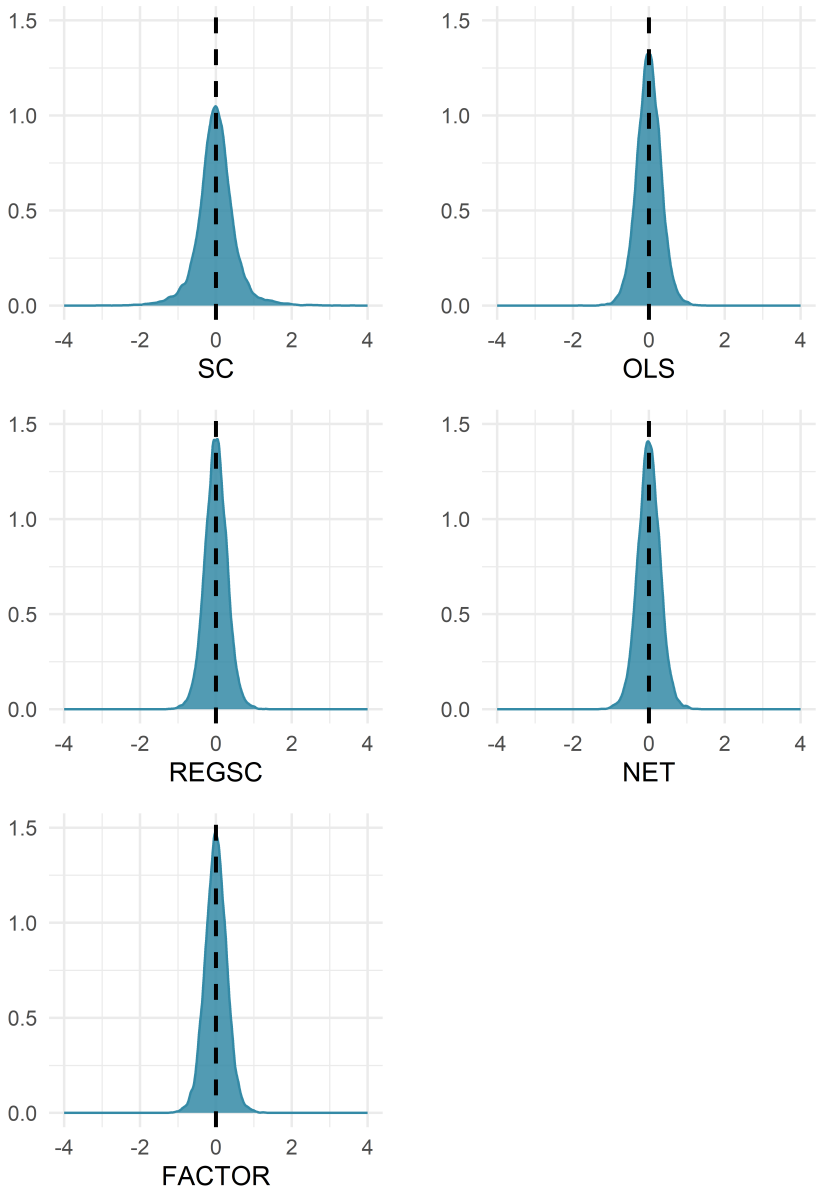

Figure 12. Bias-densities for $T_{pre} = 100$ and $T_{post} \in \{10, 20, 30\}$

Table S8. RMSFE results for $T_{pre} =\in \{20, 50, 100\}$ and $J \in \{5, 10, 15, 20, 25, 30\}$

T_{pre}	Donors	SC	OLS	REGSC	NET	FACTOR	2nd best
20	5	1.4438	1.3665	1.2988	1.2980	1.2450	NET
20	10	1.2889	1.6212	1.2294	1.2591	1.1620	REGSC
20	15	1.2444	2.4629	1.2133	1.2377	1.1266	REGSC
20	20	1.2144	NA	1.1961	1.2251	1.1132	REGSC
20	25	NA	NA	1.1976	1.2175	1.1081	REGSC
20	30	NA	NA	1.1724	1.2051	1.0901	REGSC
50	5	1.4126	1.2101	1.2038	1.1946	1.1735	NET
50	10	1.2366	1.2006	1.1318	1.1404	1.0957	REGSC
50	15	1.1888	1.2640	1.1134	1.1276	1.0740	REGSC
50	20	1.1649	1.3491	1.0999	1.1157	1.0643	REGSC
50	25	1.1515	1.4856	1.0952	1.1104	1.0570	REGSC
50	30	1.1377	1.6422	1.0867	1.1074	1.0474	REGSC
100	5	1.3886	1.1750	1.1750	1.1689	1.1580	NET
100	10	1.2205	1.1308	1.1044	1.1082	1.0835	REGSC
100	15	1.1695	1.1361	1.0810	1.0879	1.0585	REGSC
100	20	1.1371	1.1575	1.0667	1.0783	1.0443	REGSC
100	25	1.1213	1.1871	1.0580	1.0695	1.0349	REGSC
100	30	1.1105	1.2299	1.0565	1.0704	1.0346	REGSC